PETER L. BALISE

MECH. ENG. DEPT.

UNIV. OF WASH.

COMMUNICATION NETWORKS

VOL. II

THE CLASSICAL THEORY OF LONG LINES, FILTERS AND RELATED NETWORKS

BY

ERNST A. GUILLEMIN, Ph.D.

Professor of Electrical Communications
Massachusetts Institute of Technology

NEW YORK
JOHN WILEY & SONS, Inc.
London

Thirteenth printing, April, 1960

Printed in U. S. A.

PREFACE

Specifically, the present volume is intended to present a thorough treatment of the transmission line as a communication facility, leading quite naturally into the field of filter theory and its related problems. More generally, however, the discussions pertain to the field of network theory as a whole, and apply to the power as well as to the communications aspect. The communications viewpoint is chosen for the more detailed discussions because it affords at present a more suitable vehicle for illustrating the fundamental principles. This is due in part to the fact that the consideration of network functions versus frequency, which is so essential to the study of network properties, is more vital to the communications aspect. Then, too, the problem of synthesis, which has been an important motivating influence in the enlargement of our views and processes with regard to network theory, not only had its inception in the field of communications but owes its development almost wholly to workers in that field. Nevertheless, these ideas and principles are too general in nature to remain confined to one field of application, and are more recently beginning to make their appearance in the treatment also of power problems.

The content of the chapters dealing with the properties of driving-point and transfer impedances and admittances, the fundamental forms of two- and four-terminal networks in the reactive or dissipative cases, the methods of synthesis of networks from prescribed impedance or admittance functions, the subjects of network equivalence, reciprocity, duality, and the use of linear transformations should become as helpful to the power engineer as they have been indispensable to the development of the present communications aspect of network behavior and design.

The synthesis point of view in network design has been with us for hardly more than a decade. During this brief period, particularly during the latter half, the activity has been very marked, and the contributions of ideas and viewpoints have been numerous and diverse. Some of these are more fundamental and permanently useful than others. Although, at present, it is difficult to judge the ultimate utility of much of this material, an attempt has been made to select for presentation only such contributions as are thought to be of more permanent value, and to weave these into the thread of the discussions in

such a way as to obtain, as nearly as possible, a logically connected development of thought.

In the preparation of the material for this volume I have been materially aided by Dr. O. Brune, with whom I have been associated in the teaching of the subject matter on network synthesis. His many suggestions and valuable council have largely influenced the form of treatment in those parts of the manuscript dealing with the synthesis point of view. I am grateful also to Mr. H. W. Bode, of the Bell Telephone Laboratories, for his kindness in making available to me the material dealing with his contributions to the subject of filter design as well as numerous other points of view on the subject at large. With reference to the relation of the engineering formulation of the line problem to the more rigorous electromagnetic field aspects I am indebted to Dr. J. A. Stratton for helpful criticism and suggestions. To Prof. E. L. Bowles and Prof. D. C. Jackson I wish to express my appreciation for their encouragement and cooperative attitude.

E. A. G.

CAMBRIDGE, MASS.
September, 1935

CONTENTS

COMMUNICATION NETWORKS

Vol. II

CHAPTER I

THE ENGINEERING FORMULATION OF THE LONG LINE PROBLEM

1. Distinction between distributed and lumped constant systems. The transmission of intelligence by electrical means is generally accomplished in either of two ways. One usually employs a pair of parallel wires; the other relies more upon the phenomenon of wave propagation through free space or along the earth's surface. Although the second is physically more involved, and could hardly be classified under the head of a circuit problem, the two methods have much in common from the standpoint of the fundamental principles employed in their analysis. The reason for this lies in the fact that both means involve the transmission of electrical energy through continuous media. The parallel conductor problem, which will receive a large portion of our attention in this volume, is, from the engineering standpoint, closely associated with the principles of ordinary circuit theory already discussed. It is usually classified as a circuit problem of a higher order of complexity, and its treatment, therefore, is logically assumed to follow that of the lumped-constant network rather than precede it. Although this order of treatment may be justified on the basis of the mathematical aspect, it is rather illogical with regard to the fundamental physical principles involved. These have to do with that which distinguishes the networks considered heretofore from the type of network referred to as the transmission line. The latter is called a *distributed-constant* system; the former is designated as one involving *lumped constants*. Our object in the present section is to show that the distributed system is really the common ground or starting point, and that the so-called lumped-constant network is an approximation which the engineering treatment of network theory finds justified and useful. It is essential to obtain as clear a picture as possible of the basis upon which the distinction between lumped and distributed systems is made before taking up a more detailed consideration of the latter.

In the treatment of ordinary lumped networks we made an assumption which, at that time, was not specifically pointed out to the reader. For example, when we considered a battery suddenly impressed upon a coil, we assumed that this impressed force was felt simultaneously throughout the coil, and that the current which flowed in response to this force was at every instant identical in all parts of the coil. Although this assumption seems justified in view of the small physical dimensions involved, it is nevertheless not an accurate description of the phenomenon. The impressed force is not simultaneously felt throughout the coil, nor is the resulting current the same in all parts of the coil at any instant. The impressed force is propagated with a finite velocity so that various parts of the coil experience its effect at slightly different times, depending upon the rate of propagation and the physical dimensions involved. The current which flows is, therefore, also not quite the same at all points for any given instant. In general, we expect that if the coil or any combination of coils, condensers, and resistances, for that matter, are very large in size, the assumption that the response, as well as the disturbing force, is simultaneous throughout, becomes untenable. In addition to the physical size of the system, its physical nature, as well as that of the driving force, also influence the magnitude of the error involved in such an assumption. If the latter is justified, the solution of the problem is much simpler, as the reader can easily appreciate. We then say that we are dealing with a lumped system. When conditions are such that the assumption introduces too large an error, then the distributed character must be partly or wholly taken into account. Ordinarily, in the treatment of the transmission line the distributed character is only partly taken into account. The nature of this situation and the justification for the usual procedure will now be discussed.

For the engineer the above distinction between lumped and distributed systems is entirely too vague. It is necessary, therefore, that we formulate a more concrete definition, or at least a basis for such a definition. Although such a formulation seems apropos at the present, we shall have to ask the indulgence of the reader in tolerating the introduction of certain terms, which he cannot be expected to appreciate fully at this time, and which certainly will not be treated in detail until very much later in this discussion. We assume, however, that the reader has been placed in this position before, and has become somewhat accustomed to it. Unfortunately, the presentation of this subject necessitates this sort of procedure at various times.

We shall see later that any disturbance may be considered as composed of a linear superposition of simple harmonic components, which

may also be called "frequency components" since each simple harmonic function has a definite frequency whereby it may be characterized. If the disturbance is in the nature of a traveling wave of arbitrary shape, then these components are referred to as component waves. We shall also see that each component is propagated with a definite velocity, usually in the vicinity of the velocity of light. Each component wave is not only a simple harmonic function of distance but, at a fixed point, presents a simple harmonic function of time as well. This will be clear if the reader will visualize a water wave having a simple harmonic shape and traveling at a constant velocity. A cork floating on the wave and restrained from moving horizontally will be observed to bob up and down with simple harmonic motion. It should also be clear that the velocity of propagation of the harmonic wave, the length of the wave (i.e., the distance from crest to crest or trough to trough), and the frequency with which the cork bobs up and down, are not independent quantities but bear a definite relationship to one another. The cork bobs up and down once for each passage of one wave length. If we call the temporal period of the cork τ, the wave length λ, and the velocity v, then we see that:

$$\frac{\lambda}{\tau} = v = \lambda \cdot f \tag{1}$$

where f is the frequency with which the cork bobs.

Thus we find that to each "component frequency," in a given disturbance, there corresponds a definite wave length. Throughout this length the component wave changes in phase by 2π radians. The criterion as to whether or not a physical system may be considered lumped at a point may now be formulated somewhat more definitely as follows: If the shortest essential wave length in the disturbance (as determined from the highest essential frequency by the relation (1)) is very large compared with the maximum physical dimension of the system acted upon (say 50 or 100 times as large), then the maximum phase difference in this component wave at any two points in the system must be very small (not in excess of 1/50 or 1/100 of 2π radians). This is identical with saying that the component wave has the same instantaneous value at all points in the system to within a maximum error dependent upon this small phase difference. If this error is small enough to be negligible under the circumstances, then the system may be considered lumped at a point.

The rigorous application of this criterion may be rather difficult. It necessitates a knowledge of the "largest physical dimension" (which need not be the straight line from end to end of the apparatus) and also

the velocity of propagation of the shortest component wave along this path, both of which can only be estimated. However, we wish here to impress upon the reader the significance of the criterion rather than the manner of its application.

As a numerical illustration, let us consider a system which is to operate in connection with audio frequencies up to 5,000 cycles per second. Assuming that the velocity of propagation equals that of light, which is 3.10^{10} cm per sec, this corresponds to a shortest wave length of 6.10^{6} cm or 60 kilometers, which is about 37 miles. If the system consists of an arrangement of coils and condensers in a laboratory set-up or the like, it would certainly be safe to say that we may consider it lumped even though our above estimate is off by a factor of 10. On the other hand, the system may be a telephone line 50 or 100 miles long. Then we certainly could not consider the problem under the head of a lumped-constant affair. In cases which are less clear-cut than this, or where there may be more or less doubt, the criterion must be applied with more rigor, or the question may be referred to the laboratory. We might mention in this connection that in using frequencies such as occur in short-wave radio work, involving wave lengths of 10 meters or less, we are beginning to reach the limit within which our apparatus may be considered lumped. The manner in which such circuits may be effectively treated analytically still remains to be determined.

It may be well at this time to point out to the reader that where the lumped-constant treatment is no longer valid, it becomes equally difficult to adhere to our usual concepts of inductance and capacitance. For example, the inductance of a closed loop of wire is defined as the quotient of the flux linking the loop and the current flowing in the wire, assuming the wire diameter to be negligible. If the current is constant, this is a perfectly definite quantity. If the loop is excited by a very high-frequency source, however, then the current at any instant may vary considerably throughout the loop. Hence, although we could measure the total flux linking the loop at that instant, we would be at a loss as to what value of current to divide by. We could no longer assign an inductance to the loop. A similar situation applies to capacitance, which is usually defined as the ratio of the total charge carried by a pair of conductors to the potential difference between them. If the frequency of the source of excitation is very high so that the wave length is comparable with the dimensions of the conductors, then at a given instant the potential difference between them will vary considerably from point to point. In fact, we can no longer speak of a voltage between any two fixed points, because the line integral of

electric field intensity by which this potential difference is defined is no longer independent of the path of integration.

Thus we recognize that the so-called lumped-constant treatment of networks is always an approximation except in the case of steady-state behavior with constant applied forces when the laws of electro- and magneto-statics rigidly apply. Whenever the exciting forces are variable with time, as they are in almost every network problem, the electric and magnetic fields are no longer stationary, and the concepts of inductance and capacitance, which are so useful to the engineer, are no longer tenable in the light of rigid fundamental electromagnetic theory. Fortunately, however, for those frequencies with which we have to deal most commonly, the error made by disregarding this fact is negligible for practical purposes. We call such fields *quasi-stationary* to distinguish them from the very rapidly changing fields for which the error is too large to be neglected. Naturally enough, there is no sharp line of demarcation between those fields which may be considered quasi-stationary and those which can no longer be so treated. The one continuously merges with the other, and it is strictly a question of allowable tolerances as to whether the one or the other point of view may be taken.

In communication work, two general classes of problems fall in the quasi-stationary category. One is that involving the so-called lumped-constant networks, which were treated in the first volume. The other is the long transmission line problem which will occupy a large part of our attention here. The long-line problem as it presents itself in communication work may still be treated as a quasi-stationary field problem, but not as a lumped-constant problem, for reasons already pointed out. The concepts of inductance and capacitance may still be used as approximations although in a modified sense which we shall describe in more detail below. Roughly we designate such problems as the long line as distributed-constant problems to distinguish them from those where the parameters may be considered lumped. We speak of uniformly distributed inductance, capacitance, resistance, and leakage conductance as the parameters which characterize the uniform long line. The form of solution based upon such parameters is often referred to as the *engineering solution*. It carries with it a number of tacit assumptions and approximations, as may be expected from the circumstance that a rather complex system such as the long line or cable is completely described by specifying just four quantities. The justification for its introduction and use lies in the tremendous simplification which arises from this form of representation. It is safe to say that the long line as a communication channel would still be in

the experimental stage if it had not been for the development of the engineering form of solution, so great are the benefits derived from the resulting simplification. At the same time it would be a poor policy for the engineer to ignore entirely the more rigorous aspect of this very important problem as well as the process of evolution out of which the engineering formulation was born. An appreciation of this situation is useful not only for the sake of its general informational value but also because it affords a criterion whereby we may approximately determine the range of validity of the results derivable from such a method of solution. It is just as important to know the range within which a solution may be expected to be at least approximately correct as it is to determine such a solution. In fact, the solution would be quite useless or even misleading unless we could determine the corresponding limits within which it might safely be applied. Most of the following discussion is given for the purpose of elaborating upon this point.

2. Historical background regarding the theoretical treatment of the problem. Long lines were used in electrical communications for many years before any clear conception was attained regarding their theoretical behavior. The electric telegraph began to develop commercially in the early 1840's. The first attempt at an analysis of the propagation of electrical energy through a long uniform circuit was made by Sir W. Thomson (Lord Kelvin) in 1855.[1] His analysis, however, was limited in application to the single wire conductor with concentric return such as was used in undersea cable circuits. The return may actually have been through the sea water itself, but electrically this would amount to practically the same thing. The land circuits in those days were all of the single conductor with earth-return type, to which Thomson's analysis did not apply because he did not consider the inductive effect, which in such a line very materially affects the resulting behavior. His treatment takes into account only the capacitance and resistance, which, as we shall see later, is sufficient only in cable circuits and for these only over a limited frequency range.

To G. Kirchhoff must be given the credit for first introducing the inductive effect into the analysis of the long-line behavior. The results of his investigation are given in two papers published in 1857.[2] It appeared later that W. Weber[3] had independently solved the same

[1] "On the Theory of the Electric Telegraph," *Math. Phys. Papers 2,* p. 61, Proc. Royal Society, London, 1855.

[2] "Über die Bewegung der Elektrizität in Drähten," Pogg. Ann. 100, p. 193.
"Über die Bewegung der Elektrizität in Leitern," Pogg. Ann. 102, p. 529.

[3] Pogg. Ann. 100, p. 351, 1857.

problem but had withheld publication until he could include the results of certain experimental verifications which he was undertaking in collaboration with R. Kohlrausch. This duplication of effort shows that the problem must have attracted considerable attention at that time. It is also interesting to note that the results of these investigations agree in every essential detail. The premises, however, are unhappily of such an idealized nature as to make the results of little practical value. This is due to the fact that both investigators assume the single wire to be so far removed from other conductors that it may be considered isolated. For this reason the inductive effect does not appear in its true light so far as practical circuits are concerned. The single conductor with earth return cannot be considered as a single isolated conductor with any reasonable degree of rigor. Such a circuit would today be treated by reducing it to an equivalent two-conductor problem, for which the inductive effect is more easily and clearly expressible in engineering terms. The two-conductor circuit did not come into practical use, however, until the early 80's when it was found that the ground return introduced too much extraneous noise for satisfactory telephonic communication.

It may be said that the advent of the telephone into the communications field aroused a more concerted effort toward a better understanding of the theoretical behavior of long lines because of the fact that good telephonic communication requires a much more carefully designed transmission circuit than is needed for satisfactory telegraphic communication. This statement must be interpreted, of course, in terms of the speeds which were in use for telegraphic circuits in that early period. The high-speed telegraphic circuits of today require as careful a consideration of details as do any of the other facilities. At all events, this fact partially accounts for the almost thirty-year gap of relative inactivity which is evidenced in connection with the communications aspect of the long-line problem from the time of Kirchhoff's and Weber's work until the more detailed investigations of Oliver Heaviside[1] were made in 1886–87. The only other significant contribution of an earlier date was also made by Heaviside[2] in 1876 when he formulated the inductance concept in a more rigorous and practical fashion than was done by either Kirchhoff or Weber.

It must also be remembered, of course, that to Kirchhoff and Weber the Maxwellian concept of the electromagnetic field was not yet available as it was to the later investigators. Kirchhoff and Weber did not view the problem from the field standpoint at all. To them the

[1] O. Heaviside, El. Papers 2 (1886–87), p. 119 e.s. and p. 307 e.s.

[2] Phil. Mag. Vol. 1, p. 53 (Aug. 1876).

major seat of the phenomenon of propagation of electricity lay in the conductor itself rather than in the medium immediately surrounding it. Maxwell's concise formulation of the field theory enabled a much more convenient and accurate introduction of the capacitance and inductance parameters as derived quantities in terms of which an approximate representation of the problem could be carried through.

Although the field theory made possible a far better appreciation of the significant factors involved in the long-line problem, at the same time it indicated that the rigorous situation was extremely more involved than had been supposed. A partially rigorous treatment for the single-circular conductor was first made by J. J. Thomson[1] in 1886 and for the concentric cable by him[2] in 1889. A fuller treatment of the single-conductor problem was given by A. Sommerfeld[3] in 1899. The first successful attempt at an exact treatment of two infinitely long parallel wires was made by G. Mie[4] in 1900.

There have been, of course, numerous other contributors to this subject both within and following the period sketched above, but those mentioned are perhaps the most important. The significant conclusion that we can draw from a rather hasty perusal of all this work is that the problem of the propagation of electrical energy along wires is, rigorously speaking, a most complex one. None of the above investigations treat the problem fully and rigorously. For the most part an infinitely long line is assumed; the behavior is investigated only in the steady state; the nature or location of the source of energy as well as the effect of terminal conditions is left out of the picture entirely; and finally, the results are interpreted only within certain frequency limits where expansions of otherwise cumbersome expressions are possible. Some of these frequency limits are not even chosen to fit practical cases. Heaviside[5] is the only one who considers the nature of the source as well as the boundary effects both for the initial build-up or transient behavior and for the steady-state condition. He is the first, also, to consider the leakage through the insulation, in view of which the true significance of the inductance parameter may be appreciated. But his calculations consider primarily only one component wave out of a large total, namely, that one which carries the transmitted energy. His work is a first approximation only as compared with the

[1] Math. Soc. Proc. 17, p. 310, London, 1886.

[2] Roy. Soc. Proc. 46, p. 1, London, 1889. "Recent Researches in Electricity and Magnetism," Oxford, 1893, p. 262 e.s.

[3] Wied. Ann. d. Phys. 67, p. 233, 1899.

[4] Ann. d. Phys. 2, p. 201, 1900.

[5] O. Heaviside, El. Theory, Vol. 1, art. 54 e.s.; El. Theory, Vol. 2, art. 393, 394.

other, more rigorous treatments. For the engineer, however, this first approximation is usually sufficient. Heaviside's work, therefore, forms the backbone for all subsequent investigations of an engineering nature. The more rigorous attempts serve merely as a guiding background upon which further extensions as well as the degree of approximation of the present status may be based. A few of these matters will now be discussed in more detail.

3. The concept of guided waves. Before Maxwell[1] the physical picture of the propagation of electricity through a long circuit was more or less that which is frequently presented in elementary textbooks, where the hydraulic analogy to an electric circuit is given for purposes of visualization. That is, the seat of the phenomenon was taken to be within the conductor. What occurred outside the conductor could be neither definitely formulated nor described. The electrical energy was thought of as being transmitted through the conductor which, therefore, became of prime importance. In fact, if we accept this point of view altogether, it becomes impossible to conceive of a flow of electrical energy from one point to another without the aid of an intervening conductor of some sort. It has been the writer's experience that many students are quite wedded to this point of view, so much so, in fact, that to them the propagation of energy without wires (wireless transmission) becomes a thing altogether apart from other forms of transmission involving an intervening conducting medium. From this angle, the transmission by radio and that by means of a line become not only different in concept but entirely different in analytic treatment and physical interpretation.

An appreciation of Maxwell's theory of electromagnetic wave propagation brings the so-called wireless and wired forms of transmission under the same roof, so to speak. They merely appear as special cases of the same fundamental phenomenon. This is due to the fact that Maxwell makes the field the primary seat of all that takes place. The presence of a conductor merely causes the field to be broken up into various components, some of which are assigned to the conductor itself, others to the surrounding medium, and still others to the surface separating the two media. The process of analyzing a given situation consists of a simultaneous consideration of the fundamental equations which the field components must obey within the media as well as the various continuity conditions which must be met by the electric and magnetic fields at the surfaces dividing one medium from another.

It is easy to appreciate that to carry out such a line of attack on

[1] The terms "before Maxwell" and "after Maxwell" are used to refer to the periods before and after the concise formulation of Maxwell's field equations.

any but the simplest cases becomes mathematically a huge task, primarily because the method is so rigorous regarding the properties of the various media as well as the exact geometry of the configuration and the conditions at the boundaries. It is for this reason that approximations must be made, even in the so-called rigorous treatments. These approximations usually consist in neglecting certain field components relative to other, much larger ones. The outstanding difference between the usual rigorous treatment and the usual engineering treatment of the problem, however, lies in the fact that in the former the neglected terms are first demonstrated to be negligible, whereas in the latter their existence is often overlooked entirely.

If suitable approximations are made in the rigorous treatment of the usual transmission line with sending equipment at one end and an arbitrary load condition at the other, we find that, for a frequency range including the highest in practice today, only three electromagnetic wave components survive.[1] In their usual order of magnitude, starting with the largest, these are: the transmitted wave, the radial component supplying line losses, and the radiated wave which represents a loss into the surrounding space.

Of the three, the first, or the transmitted wave, is very much larger than the second, and the second is usually very much larger than the third. This last one, representing the radiated energy, has been shown to be proportional, in the steady harmonically excited state, to the square of the ratio of the distance between conductors to the line length, for the fundamental natural frequency of the line,[2] and for a flat-topped impulse, to the ratio of the distance between conductors to the length of the impulse.[3] In both, the radiated energy for practical lines is negligible as compared to that lost due to dissipation in the line itself, even at the highest frequencies. The fact that the radiation in the harmonic steady state is proportional to the ratio of wire separation to line length means that an infinite line will not radiate at all under these conditions; i.e., radiation in the steady state is due to and

[1] Other components designated by Sommerfeld as "Nebenwellen" (see "Riemann, Weber's Differentialgleichungen der Physik," F. Vieweg and S., 1927, p. 535; also D. Hondros, Ann. d. Phys. 30, 1909, p. 905) are attenuated so rapidly in metallic systems that their presence is altogether negligible. Even in dielectrics these "Nebenwellen" have been found experimentally to be extremely small compared to the main wave components (D. Hondros and P. Debye, Ann. d. Phys. 32, 1910, p. 465).

[2] For harmonics of this fundamental it is proportional to the square of the order of the harmonic.

[3] C. Manneback, "Radiation from Transmission Lines," A.I.E.E. Trans. 42, p. 289, 1923.

takes place in the vicinity of the line terminals entirely![1] This argument is hardly applicable to an infinitely long line, however, because there no steady state could ever build up, so that the phenomenon would come under the head of a transient radiation which is independent of the line ends. The result does clearly show why, for finite lengths, the radiation from a transmission line is negligible, whereas for an antenna it is very appreciable.

With the radiation from an ordinary long line negligible, there really remain only two component waves for the engineer to consider, namely, the transmitted and the dissipated waves. The first of these exists in the medium immediately surrounding the conductors, and travels parallel to their axes. Its mode of travel is determined jointly by the properties of the conductors and of the surrounding medium. The line losses are supplied out of the energy in the total wave as it passes along. This constitutes a continuous drain on the transmitted energy. This drain takes the form of a radial electromagnetic wave component which penetrates the surface of the conductor (assumed to have a circular cross-section) and is propagated toward the axis, decaying rapidly as it proceeds inward. The electric field in this component is parallel to the axis of the conductor and thus causes a current to flow in the axial direction. This current is that which is usually designated as the "transmitted" current which eventually "supplies power" to the load at the end of the line. In this formulation of the phenomenon, however, the line current appears merely as an incidental factor which causes a loss to be supplied by the major storehouse of energy residing in the surrounding medium and traveling toward the loaded end. In fact, the conductors themselves become merely the agents whose surfaces guide the surrounding wave to its ultimate goal. In order to emphasize this point of view, we speak of the transmitted wave as the **guided wave.**

The total electromagnetic wave surrounding the conductors has its surface of constant phase slightly tipped toward the conductor. The direction of propagation, which is normal to this surface, is therefore almost axial but with a slight radial inclination towards the conductors. The energy vector, which coincides with this direction, may therefore be split into an axial and a radial component. The former represents the true transmitted energy at the particular point in question; the latter represents the radial drain mentioned above. The fact that this radial component decays as it penetrates the conductor causes the accompanying electric field and hence the conductor current to become weaker nearer the axis, while it is greatest at the surface. This fact is

[1] See also G. Mie, *loc. cit.*, p. 248.

designated as the *skin effect*, which, by means of the above interpretation, becomes a logical and very easily visualized phenomenon.

Concerning the guided wave, the above physical interpretation of the transmission of electromagnetic energy brings the problem of the transmission line very much closer to that of wireless transmission. Here the antenna may really be looked upon as the source. The conductor is the earth, and the dividing surface along which the guided wave travels is the earth's surface. The geometry and the boundary conditions, as well as the properties of the media, are different, but fundamentally the problem is the same as that involving the line.[1]

4. The semi-rigorous treatment of the long-line problem. In developing the differential equations regarding the propagation of electromagnetic waves along a pair of uniform conductors of circular cross-section from the engineering aspect (which considers primarily the guided wave alone), we proceed from considerations which are approximations as compared to the rigorous formulation, but more rigorous than the final engineering form involving the inductance, resistance, capacitance, and leakage parameters. In other words, our argument proceeds from a middle ground which is neither entirely rigorous nor altogether as approximate and tacit relative to the derivation and significance of the line parameters as the engineering form. This we shall call the *semi-rigorous formulation* of our problem.

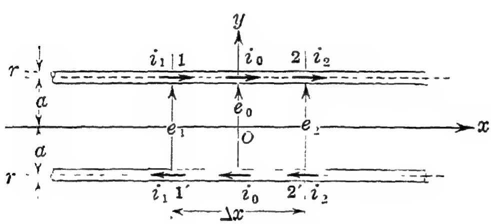


FIG. 1.—Typical internal increment of a parallel two-conductor transmission line.

The philosophy involved in this formulation is also a middle-ground philosophy, being more like the pre-Maxwellian so far as the circuital relations are concerned, but post-Maxwellian (or "post-Faraday") in the discussion and interpretation of the electromagnetic field.

To begin with, we shall consider only the presence of the fields surrounding the conductors. The fields within the conductors, and their effects upon the distribution of current throughout the conductor cross-section (skin effect), as well as the attendant effect upon the conductor resistances and internal inductances, we shall for the time being neglect.

Fig. 1 represents a typical longitudinal section of a uniform two-

[1] The ratio of the energy in the guided and radiated waves is, of course, also very different, so that the received energy may be due primarily to either one of these components or a combination of the two.

conductor line. The x-axis is laid midway between the conductor axes and parallel to these. The length Δx represents a differential length for which we wish to determine the equilibrium conditions. Since the conductors are assumed identical and symmetrically placed, the currents for a given x will be the same in both except for their directions which will be opposite. The boundaries of Δx are denoted by the cross-sections 1 and 2. The voltage between conductors at the boundary 1 is denoted by e_1, and that at the boundary 2 by e_2, while i_1 and i_2 are the corresponding currents at these points.

We now define R as the resistance per unit length of the line (both conductors in series), G as the leakage conductance (from one conductor to the other) per unit length, φ as the flux between the adjacent conductor surfaces (flux linkages not including those within the conductor cross-section) per unit length, and q as the charge per unit length carried by either conductor due to the potential difference between them. In particular, we shall write φ_0 and q_0 for the flux and charge per unit length at the point $x = 0$ where the differential length is located. In general, φ and q are functions of x, and vary from point to point for a given instant of time. Neglecting differentials of the second order, the flux linking the length Δx and the charge per conductor for this length are $\varphi_0 \Delta x$ and $q_0 \Delta x$, respectively. The loop resistance and the leakage conductance for this length are $R\Delta x$ and $G\Delta x$.

Retaining differentials of the first order only, we may, therefore, write the following equations for the voltage and current:

$$\left.\begin{aligned} e_1 - e_2 &= R \cdot \Delta x \cdot i_0 + \frac{\partial \varphi_0}{\partial t} \cdot \Delta x \\ i_1 - i_2 &= G \cdot \Delta x \cdot e_0 + \frac{\partial q_0}{\partial t} \cdot \Delta x \end{aligned}\right\}, \tag{2}$$

where the voltage terms on the left are "rises" and those on the right are "drops" in considering the loop 122′1′1. This is the same convention as to signs which is used in lumped networks. The first of these equations states that the voltage variation throughout any differential length of line is partly accounted for by resistance drop and partly by an inductive drop. The second states that the current variation is due in part to a leakage current through the insulation and in part to a charging current (displacement current) due to the electrostatic capacitance between the conductors. The values i_0 and e_0 in these equations refer to the point $x = 0$, i.e., to the location of our differential length. It is not necessary to distinguish between the values at the boundaries 1 and 2 for the terms involving these quantities because this would involve differentials of the second order.

In order to put the relations (2) in differential form, we note that by means of a series expansion

$$e = e_0 + \left(\frac{\partial e}{\partial x}\right)_0 \cdot x + \frac{1}{2}\left(\frac{\partial^2 e}{\partial x^2}\right)_0 \cdot x^2 + \frac{1}{6}\left(\frac{\partial^3 e}{\partial x^3}\right)_0 \cdot x^3 + \cdot \cdot \cdot$$

we get for $x = \pm\frac{\Delta x}{2}$, respectively,

$$\left.\begin{aligned} e_2 &= e_0 + \frac{1}{2}\left(\frac{\partial e}{\partial x}\right)_0 \cdot \Delta x + \frac{1}{8}\left(\frac{\partial^2 e}{\partial x^2}\right)_0 \cdot \overline{\Delta x}^2 + \frac{1}{48}\left(\frac{\partial^3 e}{\partial x^3}\right)_0 \cdot \overline{\Delta x}^3 + \cdot \cdot \cdot \\ e_1 &= e_0 - \frac{1}{2}\left(\frac{\partial e}{\partial x}\right)_0 \cdot \Delta x + \frac{1}{8}\left(\frac{\partial^2 e}{\partial x^2}\right)_0 \cdot \overline{\Delta x}^2 - \frac{1}{48}\left(\frac{\partial^3 e}{\partial x^3}\right)_0 \cdot \overline{\Delta x}^3 + \cdot \cdot \cdot \end{aligned}\right\}. \quad (3)$$

Writing similar expressions for current, we can form

$$\left.\begin{aligned} e_2 - e_1 &= \left(\frac{\partial e}{\partial x}\right)_0 \cdot \Delta x + \frac{1}{24}\left(\frac{\partial^3 e}{\partial x^3}\right)_0 \cdot \overline{\Delta x}^3 + \cdot \cdot \cdot \\ i_2 - i_1 &= \left(\frac{\partial i}{\partial x}\right)_0 \cdot \Delta x + \frac{1}{24}\left(\frac{\partial^3 i}{\partial x^3}\right)_0 \cdot \overline{\Delta x}^3 + \cdot \cdot \cdot \end{aligned}\right\}. \quad (4)$$

Substituting the results (4) into (2) and neglecting third order terms, we have after passing to the limit $\Delta x \to 0$

$$\left.\begin{aligned} \left(\frac{\partial e}{\partial x}\right)_0 + \frac{\partial \varphi_0}{\partial t} + Ri_0 &= 0 \\ \left(\frac{\partial i}{\partial x}\right)_0 + \frac{\partial q_0}{\partial t} + Ge_0 &= 0 \end{aligned}\right\}. \quad (5)$$

These are the differe tial equations for the long line in semi-rigorous form. They must, however, be supplemented by two further relations, since, as they stand, they involve four unknowns. For technical purposes we should like to eliminate φ_0 and q_0 so as to leave only the voltage and current to be solved for. This may be done by expressing φ_0 in terms of the line current and q_0 in terms of the line voltage. In order to carry this out we will make use of the following well-known relationships in electro- and magneto-statics. These, of course, do not rigidly apply here, but, as we shall see later, their use is justified to a close approximation except at very high frequencies.

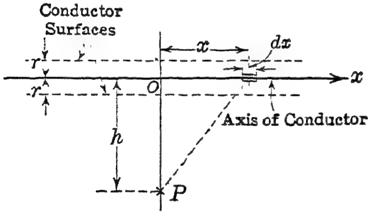


FIG. 2.—Geometrical relationships underlying equations (6) and (7).

The relations which we need here are those for the determination of the magnetic and electric fields surrounding parallel circular conductors

carrying prescribed current and charge distributions. For a single linear conductor these are usually derived assuming the current and charge to be concentrated in the conductor axis. Taking the latter as the x-axis, the magnetic and electric field intensities at a point h units perpendicularly from the axis, and located at $x = 0$ (see Fig. 2), are given by:

$$B_p = \mu \int_{-\infty}^{\infty} \frac{h \cdot i(x) \cdot dx}{(h^2 + x^2)^{3/2}} \tag{6}$$

and

$$E_p = \frac{1}{\varepsilon} \int_{-\infty}^{\infty} \frac{h \cdot q(x) \cdot dx}{(h^2 + x^2)^{3/2}}, \tag{7}$$

where $i(x)$ and $q(x)$ are functions which give the distribution of current and charge along the conductor. ε and μ are the dielectric constant and permeability, respectively, of the surrounding medium. If $q(x)$ is dissymmetrical about $x = 0$, E_p is that component perpendicular to the x-axis. These relations, of course, hold only for $h \geqslant r$, i.e., for points outside the conductor. The direction of B_p is perpendicular to the plane of the paper and its sense is determined from the direction of the current by the right-hand-screw rule, while that of E_p is in the plane of the paper, perpendicular to the x-axis, and its sense is determined by the sign of the charge q. Other than this, radial symmetry obtains.

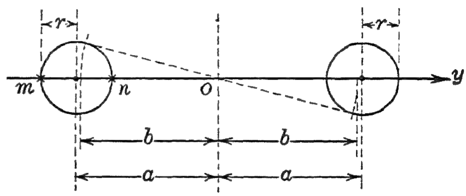


FIG. 3.—Cross-section through the uniform line of Fig. 1.

For a pair of parallel conductors as shown in Fig. 1, the situation is somewhat different. First of all, both conductors contribute toward the values of the resulting fields. In addition to this the nearness of one conductor to the other (proximity effect) also modifies matters. This modification is in general quite complex for the non-stationary state. For stationary currents and charges, however, with the assumption that the resistivity of the conductor material is negligible, the modification merely amounts to the fact that the currents and charges can no longer be considered concentrated along the conductor axes, even for the external field distribution. They can, however, be considered concentrated along axes which are eccentric with respect to the conductor axes and shifted toward the adjacent surfaces of the two

conductors.[1] This is illustrated in Fig. 3, which represents a cross-section through the line of Fig. 1. The origin of coordinates is at the center of the figure, and the x-axis is directed perpendicularly into the paper. The conductor on the left, therefore, is the lower one in Fig. 1, and the right-hand one, the upper. The conductor axes are the lines $y = \pm a$, while those axes along which the currents and charges may be considered concentrated are located by $y = \pm b$. As indicated in the figure, the distance b is determined from:

$$b^2 = a^2 - r^2. \tag{8}$$

This may be demonstrated by making use of the fact that the conductor surfaces must be equipotential loci. Since, for a uniform charge distribution along the axes $y = \pm b$, the potential is logarithmic, we get for the potential at any point in the cross-sectional plane, assuming equal and opposite charges, the difference of two logarithms or the logarithm of the ratio of the distances from this point to the axes $y = \pm b$. Constant potential loci, therefore, are such for which this ratio equals a constant. Considering two points m and n (Fig. 3) on the left-hand conductor surfaces, for example, we must have:

$$\frac{a - b + r}{a + b + r} = \frac{-a + b + r}{a + b - r},$$

from which (8) follows.

With this point in mind, we can now apply the relations (6) and (7) to the determination of the magnetic and electric field intensities at a point on the y-axis of Fig. 1, located between the adjacent surfaces of the conductors. Assuming that the top conductor carries the positive charge and the lower an equal negative one, a positive voltage will be a rise from the lower to the upper, and applying the proper rules for the directions of the electric and magnetic field intensities, we see that their vector product, which indicates the positive direction of energy flow, coincides with the positive x-direction. Making this assumption and

[1] So far as the charges are concerned, this may be seen from the fact that the opposite charges on the two conductors attract each other. The amount of the shift is determined by utilizing the fact that the conductor surfaces must be equipotential loci. Why the axes in which the currents may be considered concentrated are also shifted, and in fact are coincident with those for the charges, is not so evident. This follows, however, if the conductivity of the wires is assumed infinite. Then the skin and proximity effects, which tend to crowd the current toward the adjacent surfaces, are complete even at the lowest frequency, and may hence be assumed to be so as a limit in the case of zero frequency. This dissipationless or ideal case we shall take as our starting point for our present discussion of the external fields, and later consider the deviations from it.

using the relations illustrated in Fig. 3, the application of (6) and (7) gives:

$$B = \mu \int_{-\infty}^{\infty} \left\{ \frac{b - y}{[(b - y)^2 + x^2]^{3/2}} + \frac{b + y}{[(b + y)^2 + x^2]^{3/2}} \right\} \cdot i(x) \cdot dx \qquad (9)$$

and

$$E = -\frac{1}{\varepsilon} \int_{-\infty}^{\infty} \left\{ \frac{b - y}{[(b - y)^2 + x^2]^{3/2}} + \frac{b + y}{[(b + y)^2 + x^2]^{3/2}} \right\} \cdot q(x) \cdot dx \qquad (10)$$

where the direction of B is perpendicularly into the paper, and that of E is in the y-axis from top to bottom.

The flux per unit length and the voltage between conductors may then be found from:

$$\varphi_0 = \int_{-(a-r)}^{a-r} B \cdot dy \qquad (11)$$

and

$$e_0 = -\int_{-(a-r)}^{a-r} E \cdot dy \qquad (12)$$

respectively. Substituting (9) and (10) into (11) and (12) and carrying out the integration with respect to y, we find:

$$\varphi_0 = 2\mu \int_{-\infty}^{\infty} \left\{ \frac{1}{[(b - a + r)^2 + x^2]^{1/2}} - \frac{1}{[(b + a - r)^2 + x^2]^{1/2}} \right\} \cdot i(x) \cdot dx. \qquad (13)$$

and

$$e_0 = \frac{2}{\varepsilon} \int_{-\infty}^{\infty} \left\{ \frac{1}{[(b - a + r)^2 + x^2]^{1/2}} - \frac{1}{[(b + a - r)^2 + x^2]^{1/2}} \right\} \cdot q(x) \cdot dx. \qquad (14)$$

These relations, together with (5), completely specify the behavior of the guided wave. If the system were dissipationless, i.e., if R and G were zero, this formulation for the behavior of the guided wave would be exact, because the field distributions upon which the derivations of (13) and (14) depend are rigorously correct at moderate frequencies only under these circumstances.[1] Otherwise not only the basis for (13) and (14) but also the values of R and G will depend upon the frequency.

These relations are, however, in such a form that it would be very difficult if not impossible to determine a solution except by a method of successive approximations. This is due primarily to the fact that

[1] Here it should also be noted that (13) assumes the flux to be due to the conductor current alone. At very high frequencies the displacement current may contribute sufficiently to the magnetic field intensity to invalidate this result. For frequencies satisfying the criterion set up in the next section, however, the contribution due to the displacement current is negligibly small. Granting this for the moment, we may accept (13) as sufficiently representing the situation for our purposes.

the functions $i(x)$ and $q(x)$ are involved in the integrals (13) and (14). The engineering statement of the problem amounts to forming a first approximation toward the solution of (5), (13), and (14).

5. The engineering treatment of the long-line problem. In order to evaluate the integrals (13) and (14) approximately, we may assume that

$$\left.\begin{aligned} i(x) &= i_0 \\ q(x) &= q_0 \end{aligned}\right\} \tag{15}$$

i.e., that these functions have everywhere the same value as at the point $x = 0$, which refers to the arbitrary location of Δx in Fig. 1. This is the same as assuming that the variations of $i(x)$ and $q(x)$ in (13) and (14) affect the values of these integrals to a negligible extent. This in turn will be the case if the rates of variation of $i(x)$ and $q(x)$ are small compared to the rate of convergence of the integrals assuming $i(x)$ and $q(x)$ to be constant.

In order to investigate this, we will replace the limits in these integrals by finite values, and evaluate them under the assumptions (15). The rate of convergence of the integrals can then be studied by considering this result as a function of the finite limits. To carry this out we form the integral

$$\begin{aligned} J &= \int_{-k}^{k}\left[\frac{1}{[(b-a+r)^2+x^2]^{1/2}} - \frac{1}{[(b+a-r)^2+x^2]^{1/2}}\right]\cdot dx \\ &= 2\ln\left(\frac{b+a-r}{b-a+r}\right) - 2\ln\left(\frac{k+\sqrt{k^2+(b+a-r)^2}}{k+\sqrt{k^2+(b-a+r)^2}}\right), \end{aligned}$$

which is the integral common to both (13) and (14) after $i(x)$ and $q(x)$ are taken out. Using the relation (8) and retaining only the first significant term in an expansion of the second logarithm,[1] we find:

$$J = 2\ln\left(\frac{a+\sqrt{a^2-r^2}}{r}\right) \quad \frac{2(a-r)\sqrt{a^2-r^2}}{k^2}. \tag{16}$$

Here the first term represents the value which the integral would have with infinite limits. The second term represents the approximate error made by replacing the infinite by finite limits. Writing this as

$$J = 2\ln\left(\frac{a+\sqrt{a^2-r^2}}{r}\right)\cdot(1-\epsilon),$$

[1] The radicals are first expanded in terms of $[(b+a-r)/k]^2$ and $[(b-a+r)/k]^2$. Dropping terms above the square, the argument is simplified to $\{1 + b(a-r)/k^2\}$. Since $b(a-r) \ll k^2$, the logarithm of this quantity except for higher order terms is given by $b(a-r)/k^2$.

we find the decimal error to be:

$$\epsilon = \frac{(a - r)\sqrt{a^2 - r^2}}{k^2 \ln\left(\dfrac{a + \sqrt{a^2 - r^2}}{r}\right)}. \tag{17}$$

Noting that $a > r$, we see that, for a given k, this error becomes smaller, and hence the rate of convergence of the integral greater, for a smaller ratio of line spacing to wire diameter. For a given spacing and wire diameter, the error decreases inversely as k^2. For a prescribed error, on the other hand, the value of k may easily be determined. In terms of the latter, we may now formulate a criterion regarding the degree of approximation involved in assuming $i(x)$ and $q(x)$ constant for the evaluation of (13) and (14), namely: *If* i(x) *and* q(x) *vary so slowly as to be sensibly constant throughout the interval* $-k<x<k$, *then the error involved in assuming them constant throughout the interval* $-\infty<x<\infty$ *will not exceed the approximate value given by (17).*

In order to illustrate this numerically, suppose we consider an open telephone line with No. 10 conductors spaced 10 inches on centers. Assuming an allowable error of 5 per cent, we find by (17) that k is roughly 10 inches. Hence if the variations of current and voltage along the line (for a given instant) are slow enough so that they may be considered constant throughout a 20-inch length, then for this line the above approximation may be made with sufficient accuracy for engineering purposes.

The question as to whether $i(x)$ and $q(x)$ may be considered slowly variable depends, in the harmonically excited steady state, upon the frequency, or rather upon the wave length as determined from (1). We shall see later that, for neglected losses in the system,[1] the current and charge are simple harmonic functions of x. A wave length is that distance throughout which the functions complete one cycle. Throughout a small portion of a wave length they may be considered fairly constant. That is, if the wave length is large compared to $2k$, then the above criterion will be approximately met.

This situation is actually more favorable than the above argument would indicate, for the following reason. Any actual variations of current and charge with distance can always be approximated by the following series expansions:

$$\left.\begin{aligned} i(x) &= i_0 + i_1 \cdot x + i_2 \cdot x^2 + \cdots \\ q(x) &= q_0 + q_1 \cdot x + q_2 \cdot x^2 + \cdots \end{aligned}\right\}. \tag{18}$$

[1] With a reasonable amount of dissipation present, the situation is usually only slightly modified.

The assumptions (15) restrict these expansions to their first terms only. Now we note that the integrand in (13) and (14), except for the factors $i(x)$ and $q(x)$, is an even function of x. If (18) is substituted, and the integration carried out term by term, all terms for the odd powers in (18) vanish because their integrands are odd functions of x and hence the integrals between symmetrical limits are zero. This means that the odd terms in (18) have no effect upon the values of (13) and (14)! Consequently the current and charge may vary linearly (for example) throughout the interval $-k<x<k$ at any rate at all, and yet the result will be the same as though they were constant. The only kind of variation within this interval which disturbs the values of the integrals is a symmetrical one.

For our simple harmonic variation of $i(x)$ and $q(x)$ in the dissipationless case, this means that the vicinity of a node[1] will give rise to the same value of the integral (13) or (14) as would a constant region with the same average value. An anti-node,[2] on the other hand, is a symmetrical region. This would influence the values of the integrals. But the variation in the vicinity of an anti-node is rather slow. The cosine function, for example, varies about an average value by only 10 per cent from $-\frac{\pi}{4}$ to $+\frac{\pi}{4}$, i.e., over a quarter of a wave length in the vicinity of its maximum.

From this we may conclude that if one-quarter wave length is equal to or larger than $2k$ as determined from (17) with a reasonable value for ϵ, then the assumption as to the constancy of $i(x)$ and $q(x)$ in (13) and (14) may be made with sufficient accuracy for most engineering purposes.[3] Fixing the allowable value of ϵ at five-hundredths, this

[1] A point where the function passes through zero.

[2] A point where the function passes through its maximum.

[3] For frequencies in this vicinity or higher, the displacement current begins to have an effect upon the magnetic field intensity which determines φ_0. If we assume a non-dissipative system, equations (68*a*) and (69*a*), p. 56, next chapter, show that the displacement current, which depends upon the time derivative of the voltage, is in time phase and in space quadrature with the conduction current. Both vary sinusoidally with distance. The displacement current has a node where the conduction current has an anti-node, and vice versa. Thus, in the vicinity of a conduction current node, the displacement current is a symmetrical (even) function and hence contributes nothing to φ_0, whereas in the vicinity of an anti-node of the conduction current the displacement current is an anti-symmetrical (odd) function so that the error in φ_0 made by using i_0 for i seems partially to be compensated for by a contribution from the displacement current. Although this argument would extend the validity of the engineering assumption beyond the criterion (19), a determination of the effect of displacement current at these high frequencies requires a treatment which accounts for retardation. In this light the compensation is not

means that the engineering statement of the case, which assumes (15), is roughly justified so long as:

$$\lambda \geqslant 36\sqrt{\frac{(a-r)\sqrt{a^2-r^2}}{\ln\left(\frac{a+\sqrt{a^2-r^2}}{r}\right)}}, \tag{19}$$

where λ denotes the wave length. The equality sign gives the shortest wave length for which reasonable results may be expected from an engineering analysis. For the open line with No. 10 conductors and 10-inch spacing, this gives a smallest wave length of 78 inches. Assuming the phase velocity to be that of light, the relation (1) gives the corresponding highest frequency as $152 \cdot 10^6$ cycles per second, which is many times the highest essential frequency ever occurring on such a facility.

This argument is based upon the harmonically excited steady state. The situation is quite otherwise for transient conditions involving steep wave fronts. In the vicinity of the wave front, the above procedure with regard to the evaluation of (13) and (14) is not at all justified. It is quite obviously also not justified at points closer than k units to the line ends. How the engineering procedure is to be justified in such cases is not at all clear.[1] The fact that experimental results substantiate the engineering solution fairly well, even in the transient cases, seems to indicate that, where the criterion (19) is exceeded, errors of a compensating nature appear. It is also very likely that wave fronts so steep as to invalidate the assumption regarding reasonably constant values of i and q over the region $-k<x<k$, do not occur in any transient disturbances encountered in practice.

---

what it seems to be according to the above argument. For our purposes it is unnecessary to enter more deeply into this discussion.

[1] In an article entitled "The Guided and Radiated Energy in Wire Transmission" (A.I.E.E. J., Oct., 1924, p. 908), J. R. Carson points out that for the steady-state behavior the effect of the line ends is to introduce an infinite number of complementary waves which are ordinarily very rapidly attenuated toward the interior of the line and are, therefore, entirely negligible in comparison with those given by the usual engineering solution. No mention is made, however, as to the legitimacy of expressing these solutions in terms of inductance and capacitance parameters or the manner in which such a procedure depends upon the frequency and geometry involved. In a subsequent paper ("The Rigorous and Approximate Theories of Electrical Transmission Along Wires," B.S.T.J., Vol. 7, pp. 11–25, Jan., 1928), Carson takes up a discussion of essentially the same questions raised in the treatment given here, with no further elucidation of the effect of physical boundaries except the mention that their consideration leads to an extremely involved analysis.

The criterion (19) takes on an interesting form if we divide through by r and introduce for the ratio of wire separation to wire diameter

$$s = \frac{a}{r}.$$

Then we have

$$\frac{\lambda}{r} \geq 36\sqrt{\frac{(s-1)\sqrt{s^2-1}}{\ln(s+\sqrt{s^2-1})}}, \tag{19a}$$

or, dividing through by s,

$$\frac{\lambda}{a} \geq \frac{36}{2}\sqrt{\frac{(s-1)\sqrt{s^2-1}}{\ln(s+\sqrt{s^2-1})}}. \tag{19b}$$

These forms express the shortest wave length in terms of either the wire separation or the wire size for a given ratio between these quantities. In the above example $s = 100$, which gives

$$\frac{\lambda}{r} \geq 1556; \frac{\lambda}{a} \geq 15.56.$$

For a given ratio and wave length, the criterion imposes an upper limit upon the wire size or the spacing.

Recalling what was said earlier regarding lumped and distributed systems, the above formulation may be interpreted as characterizing the long line as a lumped system so far as its transverse dimensions are concerned, but as a distributed system in the longitudinal direction. This is evidenced by the criterion (19) which compares only the transverse dimensions with the wave length. The engineering formulation of the long-line problem is thus a peculiar hybrid combination with respect to lumped and distributed characteristics.

This fact makes it necessary to impose a restriction upon the definitions of voltage and current as they normally appear under electro- and magneto-static conditions. The voltage must be defined as the line integral of electric field intensity taken from a point on one conductor to a point on the other conductor *in the same cross-sectional plane.* Here the path of integration is restricted to lie wholly in this plane, but may otherwise be arbitrary.[1] This restriction must be imposed for the reason that the electric field may be considered conservative only in the two transverse dimensions. Longitudinally the field is non-conservative, and an arbitrary path which utilizes this dimension can lead to a different value for the line voltage. So long as the criterion (19) is fulfilled, however, the electric field is substan-

[1] This assumes an infinitely long line. If the line is finite the field at points far from the conductors may be materially influenced by the radiation components due to the end effects. The present argument leaves the latter out of consideration.

tially conservative in the transverse dimensions, so that the line voltage may still be defined in this restricted sense. A similar situation holds with regard to the line current. The latter, according to Ampère's circuital law, is given by the line integral of magnetic field intensity encircling either conductor. This integral also will not be independent of the path unless the path is restricted to lie wholly in the cross-sectional plane and unless the criterion (19) is met. The concepts of current and voltage, in terms of which the engineer prefers to characterize all network phenomena, may therefore be retained in this restricted sense for describing long-line behavior in all practical cases except those involving ultra short waves for which the criterion (19) is seriously violated.

Having thus substantiated the engineering procedure, we will continue with the formulation of the equilibrium conditions in more familiar terms. Neglecting the second term in (16), we find for (13) and (14)

$$\varphi_0 = 4\mu i_0 \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right), \tag{20}$$

and

$$e_0 = \frac{4q_0}{\varepsilon} \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right). \tag{21}$$

From these we get

$$\frac{\partial \varphi_0}{\partial t} = \frac{\partial i_0}{\partial t} \cdot 4\mu \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right), \tag{22}$$

and

$$\frac{\partial q_0}{\partial t} = \frac{\partial e_0}{\partial t} \cdot \frac{\varepsilon}{4 \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right)}. \tag{23}$$

If we now let

$$L = 4\mu \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right), \tag{24}$$

and

$$C = \frac{\varepsilon}{4 \ln\left(\frac{a + \sqrt{a^2 - r^2}}{r}\right)}, \tag{25}$$

then substitution of (22) and (23) into (5) gives

$$\left.\begin{aligned} \frac{\partial e_0}{\partial x} + L\frac{\partial i_0}{\partial t} + Ri_0 = 0 \\ \frac{\partial i_0}{\partial x} + C\frac{\partial e_0}{\partial t} + Ge_0 = 0 \end{aligned}\right\}. \tag{26}$$

These relations involve only the current and voltage, and are the desired equilibrium conditions for engineering purposes.

The quantities (24) and (25) are the distributed inductance and capacitance parameters per unit length. The product of these two is

$$LC = \varepsilon\mu, \tag{27}$$

and thus depends only upon the properties of the medium surrounding the conductors. This is to be expected since only the external fields were considered in the derivation of L and C. However, the fact that the product of L and C is independent of the geometry of the system is a rather striking result. In this connection it should be remembered that the formulae (24) and (25) are rigorously correct only in the dissipationless case, i.e., for conductor material having zero resistivity and a dielectric having zero conductivity. In the dissipative case they are very nearly correct in the vicinity of those extremely high frequencies which approach the limit indicated by (19). These matters will be discussed more fully below.

In connection with the relation (27) it should be remembered that for free space the product of ε and μ in any cgs system of units[1] equals $1/9 \cdot 10^{20}$ and has the dimensions of the reciprocal of the square of a velocity. In the practical cgs system, for example, $\varepsilon = 1/9 \cdot 10^{11}$ and $\mu = 10^{-9}$. With these values the relations (24) and (25) come out in practical henries per centimeter and practical farads per centimeter, respectively. For free space, therefore,

$$\frac{1}{\sqrt{LC}} = 3 \cdot 10^{10} \text{ cm per sec,} \tag{28}$$

which is the velocity of light. We shall see later that this is also the velocity of phase propagation of electromagnetic waves along a dissipationless open line.

6. Discussion of the line parameters. From the theoretical as well as the practical standpoint, it is important to bear in mind several matters regarding the proper determination of values for the line parameters R, L, G, and C. Since the practical considerations involved in this discussion are in some instances different from the theoretical, we shall first take up the latter, and then separately point out how they are modified by circumstances which are usually present in practice.

First we wish to point out again that the above derivations of L and C are correct only for a dissipationless system. For such a system,

[1] This refers to the electrostatic, electromagnetic, or practical systems only. In the Gaussian or mixed system of units, both ε and μ are unity, but there a factor c^2 (the square of the velocity of light) enters into the derivation so as to give rise to (28) just the same.

the field distributions are entirely outside of the conductors at all frequencies, and are correctly given by considering the charges and currents concentrated along the axes $y = \pm b$ as shown in Fig. 3. This is true up to frequencies at which the displacement current parallel to the conductor axes becomes appreciable, beyond which the engineering formulation can no longer be used without modification. The dissipationless case is therefore the ideal one in which the bothersome variation of the line parameters with frequency is entirely absent.

As soon as the resistivity of the conductor material is considered, the situation becomes markedly different, particularly with respect to the inductance parameter. The capacitance parameter varies so little with frequency, even with an appreciable amount of conductor resistance, that it may be considered constant and given by (25) throughout the practical frequency range. This is due primarily to the fact that the charges may still be considered concentrated along the axes $y = \pm b$ in Fig. 3, even when resistance is present, so that the electric field distribution remains unchanged and hence the same procedure in the derivation of C applies. This is not so in the case of the inductance. As soon as the slightest amount of resistance is present, the skin and proximity effects for zero frequency, or for frequencies very close to zero, are entirely absent instead of being substantially complete. That is to say, the current distribution throughout the conductor cross-sections is uniform for zero frequency as well as almost uniform for very small frequencies. Hence the conductor currents must be considered concentrated along the axes $y = \pm a$ instead of along $y = \pm b$ so far as the external magnetic field distribution is concerned. This means that in the derivation of the inductance parameter we must replace b by a in the integral (9). If we do this, we obtain instead of (24)

$$L = 4\mu \ln\left(\frac{2a - r}{r}\right). \tag{24a}$$

This is the correct value for the inductance due to the external field in the stationary state (zero frequency) for the dissipative case. Here the current is uniformly distributed throughout the conductor cross-sections, so that the magnetic field penetrates into the interior of the conductors instead of being altogether external as in the dissipationless case. Hence (24a) must be supplemented by that contribution to the total inductance due to the internal flux linkages. This amount is found to be[1]

$$L_i = \mu_i \tag{24b}$$

[1] See for example L. F. Woodruff, "Principles of Electric Power Transmission and Distribution," John Wiley & Sons, 1925, p. 17, e.s.

for the stationary state, μ_i being the permeability of the conductor material. The total inductance for zero frequency is the sum of (24a) and (24b) as soon as any amount of resistance is present.

As the frequency is increased gradually from zero to the upper limit discussed above, the value of the total inductance parameter varies continuously from its constant-current value to the value given by (24). The latter, which is the dissipationless value at all frequencies, is correct for the general dissipative case only at the high-frequency limit where skin and proximity effects are complete so that all the flux linkages are external, and the external field distribution is given by the picture which was had for the non-dissipative case. This means that, as the frequency increases from zero, the internal portion (24b) gradually disappears, and the external portion (24a) gradually merges with the value (24). If the conductor resistance is small, this variation proceeds very rapidly as the frequency increases. In fact, if the resistance is considered extremely small, the transition from the sum of (24a) and (24b) to the value (24) is practically complete even for very small frequencies. For vanishing conductor resistance, it is complete from the start. This is the line of reasoning which justifies (24) in the dissipationless case even for zero frequency, which is there looked upon as a vanishingly small frequency.

Since the value (24a) is smaller than (24), we see that the external inductance increases with frequency, the percentage increase depending upon the geometry of the configuration, i.e., upon the distance a as compared to r. The internal inductance decreases as the frequency increases. How the net inductance varies throughout the frequency range is quite a complex problem, and depends upon the geometry. If the spacing of the conductors is large compared to their radii, the difference between (24a) and (24) is inappreciable, and the net change is almost wholly attributable to that of (24b), which is a decrease with frequency. If the conductors are closely spaced, as in cables, (24) may be considerably larger than (24a), and the net change may be an increase.

The exact theoretical variation of the resistance parameter with frequency is due to changes in the magnitude and distribution of internal flux linkages. The net effect is the result of two tendencies, one of which is to crowd the current toward the surface in all radial directions, and the other to increase the current density in the adjacent conductor surfaces. Although the first of these (the skin effect) tends toward an infinite surface density as the frequency becomes infinite, the second (proximity effect) merely tends toward a maximum one-sided circumferential distribution which still leaves some current at all

points on the conductor contour. The skin effect, of course, increases the conductor resistance materially with frequency, and tends to make it infinite for an infinite frequency. The superposed proximity effect, which becomes more pronounced for a smaller ratio of separation to conductor radius, may greatly increase the resistance over that due to the normal skin effect at a given frequency. In open telephone lines, the ratio a/r is large, and the proximity effect almost negligible. In cables, however, it may cause the resistance to increase with frequency many times faster than it would from the skin effect alone.

Finally the variation of the leakage parameter with frequency may also be theoretically determined on the assumption of a constant conductivity of the dielectric, which is supposed to surround the conductors uniformly and extend indefinitely in the transverse directions. Under these conditions, the change in the leakage parameter is due solely to changes in the external field distributions which, as we have seen, are very slight for all frequencies up to the limit (19). The calculated variation in the leakage parameter is thus negligible.

Theoretically, therefore, the parameters G and C may be considered constant for all frequencies for which the engineering formulation of our problem is valid. Of the parameters R and L, the first is the more variable, particularly where the ratio a/r is small. Starting with its d-c value, this parameter tends toward infinity as the frequency increases without limit. The inductance may increase or decrease with frequency depending upon the ratio a/r. In any event the variation tends toward the dissipationless value at the upper frequency limit, and hence the variation is never comparable with that of R. For large a/r ratios the variation in the total inductance is due almost entirely to the internal portion, which in such cases is a small fraction of the total, so that the entire variation may be negligible.

In actual telephone lines and cables, the variation of R with frequency is found to agree, within experimental error, with the theoretically predicted variation.[1] This variation in the usual open telephone lines employing copper conductors may run from 250 to 400 per cent over a frequency range of 500 to 50,000 cycles, the larger variation occurring for the larger conductor sizes. Owing to the relatively large a/r ratio, this variation in open lines is due almost entirely to skin effect. In lines employing iron wire, the variation is considerably larger because of the higher internal permeability. In cables the *percentage* variation in R, although contributed to by the proximity effect, is considerably

[1] E. I. Green, "Transmission Characteristics of Open-Wire Telephone Lines," B.S.T.J., Oct., 1930.

less on account of the smaller wire size and consequent higher initial resistance.

The variations in the parameters L and C for practical circuits also agree substantially with theoretical predictions. For open lines the inductance decreases roughly from 2 to 4 per cent over a frequency range from zero to 50,000 cycles, the larger variation occurring for the larger conductor sizes and smaller spacings. The capacitance parameter, which theoretically should not vary by any measurable amount, does increase from 0.5 to 4 per cent over the same frequency range in open wire lines. This is due to the fact that the theory does not take into account the effect of insulators, pins, and cross-arms which are necessary in the support of open lines. The use of metal pins, or pins with metallic thimbles, introduces additional lumped capacitance at the supporting points, particularly when pairs of pins or thimbles are bonded. When numerous leakage paths over the surfaces as well as through the walls of the insulators are taken into account, these lumped capacitances may be approximately attributed to the resulting capacitance effect of a ladder network consisting of small condensers and large leakage resistances.[1] This rather complicated structure presents a capacitance component which has a rising frequency characteristic. It is this that accounts for the small observed increase in the net capacitance parameter on such facilities.

The widest difference between theoretical and observed variations is found in the leakage parameter. This, however, is to be expected since the theoretically assumed constancy of the dielectric conductivity is quite far from the truth. Whereas the theory predicts no observable variation in the leakage conductance, it is found both in open telephone lines and in cables to be *almost a linear function of frequency* over a range of 1000 to 50,000 cycles.[2] For open lines, in dry weather, this variation is roughly 5000 per cent of the low-frequency value, and for cables it is more than 8000 per cent! In addition to this, the wet-weather values for open lines are from three to twelve times the dry-weather values, the larger differences being observed at the low frequencies.

It may seem to the reader that, in view of such variations in the line parameters, the long transmission line could hardly be treated analytically according to methods which rigorously apply only to constant-parameter networks. The variation in the leakage conductance seems particularly discouraging from this point of view. But, as we shall see later, the effect of the leakage parameter upon the per-

[1] L. T. Wilson, "A Study of Telephone Line Insulators," B.S.T.J., Oct., 1930.

[2] E. I. Green. *loc. cit.*

formance of open lines and cables is slight if not altogether negligible over the usual frequency ranges for which such facilities are used. The resistance, which is the only other parameter whose variation is appreciable, may alone be considered as seriously affecting this question. At a given frequency, however, the parameters are constant. The solution for a single frequency will, therefore, be correct provided the proper parameter values for that frequency are used. Obviously the superposition of a finite number of frequencies may be treated in this manner also. A real difficulty arises in the treatment of the transient behavior, for which average values of the parameters will have to be selected unless this situation is considered in the Fourier sense, whereby the problem is again expressed in terms of the harmonic steady state. These matters, however, will be more fully appreciated later on. In the meantime it is quite evident that a treatment which assumes the constancy of the line parameters, in spite of what actual cases may seem to indicate, will have to be the point of departure in all subsequent modifications. The following chapters, therefore, are devoted to a thorough study of such a theoretically ideal line.

7. Bibliography. In the above discussion our object has been twofold. First, to acquaint the reader somewhat with the basis upon which the engineering considerations of the long-line problem rest so that he will have a better appreciation of their limitations; and second, to arouse his interest in pursuing this problem from a more rigorous angle at a later date when he shall have become thoroughly familiar with the elementary treatment. Such a further study in the region beyond which the present formulation may be applied certainly seems justified in view of the increasing tendency toward the use of higher frequencies in communication work, particularly in the field of radio. This applies equally to networks which at present are considered lumped as it does to uniform systems of larger dimensions.[1] Such further study will, of course, require a more thorough knowledge of the fundamental concepts involved than may be gained from the present rather brief survey; hence the literature, both early and recent, which bears upon this subject will have to be consulted. Rather than append a lengthy bibliography here, we refer the reader for an earlier survey (up to 1906) to a summary by M. Abraham[2] where many additional references will be found. For a more recent summary with bibliographic notes, the article

[1] For a general discussion of this situation see J. R. Carson's paper, entitled "Electromagnetic Theory and the Foundations of Electric Circuit Theory," B.S.T.J., Vol. 6, pp. 1–17, January, 1927.

[2] M. Abraham, "Elektromagnetische Wellen," Encyclopädie der mathematischen Wissenschaften, Vol. V^2, pp. 514–538.

by C. Manneback already referred to will be found useful. A rather comprehensive treatment, particularly with regard to the theoretical variation of line parameters with frequency, will be found in a Bureau of Standards publication by C. Snow.[1]

[1] Chester Snow, "Alternating Current Distribution in Cylindrical Conductors," Scientific Papers of the Bureau of Standards, Vol. 20, pp. 277–338, 1925; or paper S 509.

PROBLEMS TO CHAPTER I

1–1. A two-wire transmission line which is to be used as a feeder for an antenna has a ratio of wire separation to wire diameter of 100. The resistance and leakage may be neglected.

(*a*) What are the distributed inductance and capacitance parameters in practical henries and farads per meter?

(*b*) If a frequency corresponding to a wave length of 30 meters is to be transmitted, what is the maximum wire separation for which an engineering analysis may be expected to be reasonably correct?

1–2. Consider a two-wire transmission line whose conductor axes are spaced d centimeters apart and whose common radius is r centimeters with $r \ll d$. The conductors are assumed to carry uniform charge distributions of q_1 and q_2 coulombs per unit length. Show that the respective conductor potentials e_1 and e_2 may be expressed in terms of the charges by means of the linear relations

$$\begin{aligned} e_1 &= s_{11}\, q_1 + s_{12}\, q_2 \\ e_2 &= s_{21}\, q_1 + s_{22}\, q_2 \end{aligned}$$

in which the coefficients (called the partial elastance coefficients) are given by

$$s_{11} = s_{22} = \frac{2}{\varepsilon} \ln \frac{1}{r};\ s_{12} = s_{21} = \frac{2}{\varepsilon} \ln \frac{1}{d}.$$

These logarithmic potentials are understood to be determined to within a common additive constant and are referred to a common datum located at a distance X centimeters from one of the conductors in the limiting sense $X \to \infty$. Show that $q_2 = -q_1$ follows from $e_2 = -e_1$, and that the capacitance parameter per unit length is then given by

$$C = \frac{q_1}{e_1 - e_2} = \frac{\varepsilon}{4 \ln \dfrac{d}{r}}.$$

1–3. Assume the conductors of Problem 1–2 to carry longitudinally uniform currents of i_1 and i_2 amperes in coincident reference directions. If v_1 and v_2 denote the voltage drops per unit length along the respective conductors (both in the same reference direction), show that for negligible conductor resistance these may be expressed by the relations

$$\begin{aligned} v_1 &= l_{11}\, i_1' + l_{12}\, i_2' \\ v_2 &= l_{21}\, i_1' + l_{22}\, i_2' \end{aligned}$$

in which the coefficients (called the partial inductance coefficients) are given by

$$l_{11} = l_{22} = 2\mu \ln \frac{1}{r};\ l_{12} = l_{21} = 2\mu \ln \frac{1}{d},$$

and the primes on the currents indicate differentiation with respect to time. In a

manner similar to the logarithmic potentials in Problem 1–2, the partial flux linkages $\varphi_{rs} = l_{rs}i_s$ are determined to within a common additive constant and are referred to a common datum in the same limiting sense. Show that $i_2' = -i_1'$ follows from $v_2 = -v_1$ and for sinusoidal currents $i_2 = -i_1$. Under these conditions derive the inductance parameter per unit length and show that it is given by

$$L = 4\,\mu \ln \frac{d}{r}.$$

1–4. Consider a longitudinally uniform transmission line consisting of six identical conductors. The cross-sectional pattern forms a rectangular grid of which the groups of three conductors each lying in the same longitudinal plane form the down and return paths respectively for the net current which divides between the conductors in a group. The charges per unit length (all assumed positive) and the conductor currents (all in coincident reference directions) are denoted by $q_1, q_2, \ldots q_6$ and $i_1, i_2 \ldots i_6$, respectively. The common conductor radius is r, and the perpendicular distances between conductor axes are d_{rs}, $(r, s = 1, 2, \ldots 6)$ where the subscripts refer to the similarly numbered conductors. Again assume $r \ll d_{rs}$. Show that if we define the set of partial elastance coefficients

$$s_{rs} = \frac{2}{\varepsilon} \ln \frac{1}{d_{rs}}, \text{ with } d_{rr} = r,$$

and the set of partial inductance coefficients

$$l_{rs} = 2\,\mu \ln \frac{1}{d_{rs}}, \text{ with } d_{rr} = r,$$

the following systems of linear equations are obtained for the conductor potentials $e_1, \ldots e_6$ and voltage drops per unit length $v_1, \ldots v_6$ (neglecting conductor resistances)

$$\sum_{s=1}^{6} s_{rs}\, q_s = e_r; \quad (r = 1, 2, \ldots 6), \tag{a}$$

and

$$\sum_{s=1}^{6} l_{rs}\, i_s' = v_r; \quad (r = 1, 2, \ldots 6). \tag{b}$$

Show that the longitudinal uniformity demands that

$$e_1 = e_2 = e_3 = -e_4 = -e_5 = -e_6 = e$$
$$v_1 = v_2 = v_3 = -v_4 = -v_5 = -v_6 = v$$

and from this total of relations find the division of the charges and currents (assumed sinusoidal) between the conductors in each group, noting that

$$q = q_1 + q_2 + q_3$$
$$i = i_1 + i_2 + i_3$$

are the total charge and current per group but that these do *not* divide uniformly (skin and proximity effects for the conductor groups). Show that the division of charges is the same as that for the currents, and that this is due to having neglected the conductor resistances. Finally, derive expressions (using determinant notation) for the inductance and capacitance parameters based upon the relations

$$L = \frac{2v}{i'} \text{ and } C = \frac{q}{2e};$$

and show that in this non-dissipative case

$$LC = \varepsilon\mu.$$

1–5. For Problem 1–4 show that the consideration of the conductor resistances leads to the same problem formulation with the one difference that in the system of equations (*b*) the terms $l_{rr}i_r'$ on the principal diagonal are replaced by $(\rho_r i_r + l_{rr}i_r')$ where ρ_r are the conductor resistances per unit length. Show that the division of currents is now no longer identical with that for the charges and as a consequence $LC \neq \varepsilon\mu$.

1–6. Generalize the conductor configuration of Problems 1–4 and 1–5 by considering a total of n conductors arbitrarily spaced in the cross-sectional plane, of which the first m conductors together form the down path and the remaining n-m the return where m is not necessarily equal to $n/2$. First neglect conductor resistance and show that the group skin and proximity effects are the same for charges and currents (assumed sinusoidal) and that then $LC = \varepsilon\mu$. Second, consider conductor resistances and show that this is in general no longer true but holds in the limit of very large frequencies. On the basis of these results prove the correctness of the statement in the text that the skin and proximity effects for a pair of large conductors is the same for charge and current when dissipation is neglected and that this is approximately true in the dissipative case for relatively high frequencies. Derive a criterion for the term "relatively high."

1–7. Apply the results of the preceding problems to numerical cases and various simple cross-sectional conductor patterns (grid or regular polygon arrangements) as assigned by the instructor or chosen by the individual. In each case determine the product LC and the quotient L/C (the latter determines the characteristic impedance of the line as discussed in the succeeding chapters).

CHAPTER II

THE STEADY-STATE SOLUTION TO THE LONG-LINE EQUATIONS

1. The general linear character of the equations. In this chapter we wish to discuss primarily the method of obtaining the steady-state behavior of the long line when it is excited by means of a simple harmonic driving force. This we shall do on the basis of the engineering formulation of the equilibrium conditions discussed in the previous chapter, and expressed by the equations (26), which we repeat here for convenience, omitting the subscripts "naught" since the point $x = 0$ may be considered arbitrarily located

$$\left.\begin{aligned} \frac{\partial e}{\partial x} + L\frac{\partial i}{\partial t} + Ri = 0 \\ \frac{\partial i}{\partial x} + C\frac{\partial e}{\partial t} + Ge = 0 \end{aligned}\right\}. \tag{29}$$

There is one outstanding characteristic which these equations have in common with those for lumped-constant networks, namely, that the individual terms are all linear in e and i. This linearity of the equations is a valuable property from the standpoint of the ease of obtaining solutions when several forces are impressed either simultaneously or in succession. In fact, the validity of the superposition principle, the usefulness of which has been appreciated in the treatment of lumped-constant networks, depends upon the linearity of the system (29). It is well to bear in mind, however, that the equilibrium conditions are linear only in so far as the approximations involved in the engineering formulation are justified, because the linearity would obviously be destroyed as soon as the functional dependence of L or C upon the current or voltage were taken into account.

It may also be well to point out to the student an essential difference between the general form of the equilibrium conditions (29) and those for the usual lumped system. In the latter the currents and voltages are functions of time only, their location in the various meshes of the system being indicated by suitable subscripts. The equilibrium conditions there are given by as many simultaneous equations as there are independent meshes. Here the system is not broken up into a discrete number of meshes but is in every respect similar to a continuous medium.

The outstanding characteristic of such a medium is that the dependent variables (in our case e and i) are continuous functions of the point of location within the medium instead of being discontinuous as are, for example, the currents as we go from mesh to mesh in a lumped network. In the present instance the system is continuous only with respect to one dimension, namely, x, which is the distance along the line. This is so because we have described the condition of the system wholly in terms of the voltage and current functions and have thus, formally at least, suppressed the general space-wave character of the phenomenon.

This continuous or uniformly distributed character of the system brings with it at once an inherent simplification as well as an apparent complication in the expression of its equilibrium condition. In the light of the lumped-network procedure, we would here be faced with a system with virtually an infinite number of meshes of infinitesimal size, and a description of its dynamic equilibrium would therefore entail an infinite number of "mesh equations." The fact that the system is uniform, except for its boundaries (the ends of the line), makes it possible to avoid much needless writing of equations, for all those for any internal infinitesimal mesh are identical in form. It is therefore only necessary to express the equilibrium conditions for a single infinitesimal or differential internal mesh, as is done by the equations (29). This is an obvious simplification in form. The complication lies in the fact that we are mathematically faced with a novel feature in our differential equations. First of all, the variables e and i are functions of the two quantities x and t; hence partial derivatives appear for the reason that distance and time are involved simultaneously, and in order to consider their effects separately the partial derivative scheme is resorted to. Secondly, the question as to whether the system is being driven or is force-free is not contained in these equations. They are homogeneous in either case. In the lumped system the equations were homogeneous only in the force-free case. The excited condition led to a system of inhomogeneous equations. This situation, which is a bit puzzling at first, is due to the fact that the equations (29) really do not state the equilibrium conditions of the entire system but only of the internal portion, i.e., exclusive of the line ends at which the exciting force takes hold. It is, therefore, natural that a description of the mode of excitation should not appear in these equations. Mathematically this means that the effect of the presence or absence of excitation at the line ends will subsequently have to be taken into account, i.e., after formal solutions to the system (29) have been found. This latter process is spoken of as taking the effect of *boundary conditions* into account. For the lumped system we are already familiar with a

similar process regarding initial conditions which are in a sense the temporal boundaries about which the differential equations contain no information. Our present problem, involving partial instead of total differential equations, is therefore complicated only by the fact that boundary as well as initial conditions must be subsequently met by the formal solutions, and it is only in the boundary conditions that the effect of an exciting force will be felt.

The question of meeting initial conditions arises, of course, only in connection with the determination of the transient portion of the solution and hence only when we are interested in the complete solution to a given problem. This part of the picture is the same as for the lumped-constant problem and needs no further comment here. In the present chapter we are interested only in obtaining the steady-state solution for harmonic excitation, and hence the question as to initial conditions does not enter into the discussion. In the following we shall give first a general mode of procedure which may be followed for obtaining either the forced or the force-free behavior, and then discuss a shorter process of integration particularly adapted to obtaining the harmonically excited steady state alone.

2. Formal solution by separation of variables. In attempting to find solutions to the system (29), it might seem obvious that the first step is to separate e and i, i.e., obtain separate equations for the voltage and current. Although this procedure is perfectly good, and we shall begin by adopting it, it is not necessarily the most direct method of approach as will be seen later.

The elimination of i may be accomplished as follows: Let the first equation (29) be differentiated partially with respect to x and the second with respect to t so as to get

$$\frac{\partial^2 e}{\partial x^2} + L\frac{\partial^2 i}{\partial t\,\partial x} + R\frac{\partial i}{\partial x} = 0, \tag{30}$$

$$\frac{\partial^2 i}{\partial x\,\partial t} + C\frac{\partial^2 e}{\partial t^2} + G\frac{\partial e}{\partial t} = 0. \tag{31}$$

Now since

$$\frac{\partial^2 i}{\partial x\,\partial t} = \frac{\partial^2 i}{\partial t\,\partial x},$$

we obtain from (31)

$$\frac{\partial^2 i}{\partial t\,\partial x} = -C\frac{\partial^2 e}{\partial t^2} - G\frac{\partial e}{\partial t}, \tag{32}$$

while from the second equation (29) we see that

$$\frac{\partial i}{\partial x} = -C\frac{\partial e}{\partial t} - Ge. \tag{33}$$

Using these in (30) we get

$$\frac{\partial^2 e}{\partial x^2} = LC\frac{\partial^2 e}{\partial t^2} + (RC + LG)\frac{\partial e}{\partial t} + RGe, \tag{34}$$

which is the desired equation in e alone.

In a similar manner, by differentiating the first equation (29) partially with respect to t and the second with respect to x, we obtain an identical equation for the current, namely

$$\frac{\partial^2 i}{\partial x^2} = LC\frac{\partial^2 i}{\partial t^2} + (RC + LG)\frac{\partial i}{\partial t} + RGi. \tag{35}$$

The equations (34) and (35) are called *wave equations* for the reason that the solutions for e and i have the character of traveling waves, as will be seen later.

The next step in the process of solution involves the separation of the independent variables x and t which are simultaneously involved. This separation, in wave equations of this type, can always be done by assuming a form of solution which is quite logical on the basis of our experience with lumped networks. There we found that all the currents and voltages are equal to the same function of time with the exception that different amplitudes and phases are involved in the different meshes. Therefore, it is logical to assume here that the voltage, for example, is given by a time function whose amplitude is a function of x only. The same argument holds with respect to the current. This is the same as to say that the voltage or current ought to be given by the product of a function of distance and a function of time.

In order to illustrate the generality of the method, let us assume for the voltage equation (34) that

$$e(x, t) = g(x) \cdot h(t). \tag{36}$$

Indicating differentiation by means of primes, substitution into (34) gives

$$g''h = LCgh'' + (RC + LG)gh' + RGgh. \tag{37}$$

Dividing through by gh this becomes

$$\frac{g''}{g} = LC\frac{h''}{h} + (RC + LG)\frac{h'}{h} + RG. \tag{38}$$

Here we see that the left hand side is a function of x only while the right-hand side is a function of t only. But a function of x can equal

a function of t for all values of x and t only when each reduces to the same constant. Denoting this constant by α^2, we therefore have

$$g'' - \alpha^2 g = 0, \tag{39}$$

and

$$h'' + \left(\frac{R}{L} + \frac{G}{C}\right)h' + \left(\frac{RG - \alpha^2}{LC}\right)h = 0. \tag{40}$$

These are now total differential equations in x and t respectively. They are linear equations with constant coefficients of the general type with which we are already familiar from our experience with lumped networks. Hence we know without further comment that they are satisfied by exponential functions, as may be verified through substitution. This is nothing surprising so far as the time function is concerned. Indeed, we should have suspected this from the beginning, but we wished here merely to show that the above process of separating the variables x and t is perfectly general and has nothing to do with the fact that the solutions are of the exponential form. That is why we chose to denote the distance and time functions by means of the noncommittal letters g and h, respectively.

Knowing that the time function is an exponential, we can arrive at a result more quickly by making this assumption directly in (34) and (35). To show how this proceeds let us assume

$$\left.\begin{aligned} e(x,t) &= E(x) \cdot e^{pt} \\ i(x,t) &= I(x) \cdot e^{pt} \end{aligned}\right\}. \tag{41}$$

These are really the same kind of assumptions as were made for lumped networks, except that here the amplitudes E and I are not constants but continuous functions of the distance along the line. Our heuristic method of procedure would therefore have suggested these forms from the beginning. Substitution into the wave equations (34) and (35) gives, after rearrangement of terms and cancelation of the time function e^{pt},

$$\left.\begin{aligned} \frac{d^2E}{dx^2} - \alpha^2 E &= 0 \\ \frac{d^2I}{dx^2} - \alpha^2 I &= 0 \end{aligned}\right\}, \tag{42}$$

where we have put

$$\alpha^2 = (R + Lp)(G + Cp). \tag{43}$$

Hence the assumptions (41) will be valid and useful provided the distance functions E and I can be determined so as to satisfy the differ-

ential equations (42) as well as any special conditions that we wish to impose at the boundaries.

The equations (42) are obviously satisfied by exponentials. For example, let us assume

$$E(x) = Ae^{mx}. \tag{44}$$

Then

$$\frac{d^2E}{dx^2} = m^2Ae^{mx} = m^2E,$$

so that the first equation (42) becomes

$$(m^2 - \alpha^2)E = 0,$$

which demands that for a non-vanishing E we must have

$$m = \pm \alpha.$$

Hence the complete expression for E is given by the sum of two such terms as (44), one for each value of m, namely,

$$E(x) = A_1e^{-\alpha x} + A_2e^{\alpha x}. \tag{45}$$

Similarly we get

$$I(x) = B_1e^{-\alpha x} + B_2e^{\alpha x}. \tag{46}$$

These are the formal expressions for the distance functions.

We are now ready to take into account the fact that the system is being driven by a harmonic force at one end of the line. Let us assume here a generator having the following terminal voltage function

$$e_g = \mathcal{R}e\,[E_ge^{j\omega t}], \tag{47}$$

where ω is the angular frequency, and the $\mathcal{R}$ sign has the usual significance of taking the real part of whatever it operates upon, just as in the treatment of lumped networks. This sign may, of course, be dropped temporarily and later inserted before interpreting the instantaneous voltage and current functions.

We now make use of our experience with linear systems by saying that, when such a system is driven by a harmonic force at any point, then the steady-state voltage and current response at all points will be harmonic and of the same frequency as the driving force. This argument makes it clear that in the expressions (41), we must put $p = j\omega$ and thus write

$$\left.\begin{aligned} e(x, t) &= \mathcal{R}e\,[E(x)\cdot e^{j\omega t}] \\ i(x, t) &= \mathcal{R}e\,[I(x)\cdot e^{j\omega t}] \end{aligned}\right\}. \tag{48}$$

These expressions, together with (45) and (46), constitute the complete formal solutions to the harmonically excited steady state.

3. Direct solution for the steady state. The above procedure for obtaining the steady-state solutions is perhaps more lengthy and explicit than necessary, but we have given it in the hope that it might afford the reader a more thorough insight into the mechanism involved. Then too, owing to its more general mode of approach, it will yield the transient or force-free behavior as well as the steady-state which is our more immediate objective. In a later chapter we shall discuss the transient solution in detail, and will find that the forms (41) hold there also except that the values for p, instead of being equal to the value $j\omega$, are the natural modes of the system—again just as in lumped networks.

However, if we are interested only in a harmonic steady state, and have given a driving force of the form (47), then our experience should at once tell us that the forms (48) are reasonable, and that we ought to substitute these directly into the given equations (29) without bothering to separate e and i. Dropping the $\Re e$ sign, these substitutions immediately yield

$$\left.\begin{aligned} \frac{dE}{dx} + (R + jL\omega)I &= 0 \\ \frac{dI}{dx} + (G + jC\omega)E &= 0 \end{aligned}\right\}, \tag{49}$$

where the factor $e^{j\omega t}$ has been canceled. Here E or I may easily be eliminated by differentiating either equation with respect to x and substituting for the single derivative term from the other equation. The reader will thus see that we obtain

$$\left.\begin{aligned} \frac{d^2E}{dx^2} - \alpha^2 E &= 0 \\ \frac{d^2I}{dx^2} - \alpha^2 I &= 0 \end{aligned}\right\}, \tag{42a}$$

with

$$\alpha^2 = (R + jL\omega)(G + jC\omega), \tag{43a}$$

which is the same as (43) with $p = j\omega$. The pair (42a) is identical with (42), and hence the solutions (45) and (46) follow as before.

4. Boundary conditions and the evaluation of integration constants. The next step is to adapt the formal solutions (45) and (46) to the boundary conditions of our problem. For this purpose we have available the integration constants A_1, A_2, B_1, and B_2. Here we immediately notice an apparent defect, however. Our line has obviously only two ends, at each of which one relationship between voltage, current, and terminal impedance may be set up. This process will therefore suffice to determine only two constants, whereas apparently we have four. Obviously only two of these can be arbitrary. This we recognize to

be so if we realize that the relations (49) definitely link the voltage and current amplitudes. This fact we have not yet made use of but shall proceed to do so immediately.

Substituting (45) and (46) into (49), and writing for brevity

$$\left.\begin{aligned} y &= G + jC\omega \\ z &= R + jL\omega \end{aligned}\right\}, \tag{50}$$

we find after grouping terms with $e^{-\alpha x}$ and $e^{\alpha x}$

$$\left.\begin{aligned} (zB_1 - \alpha A_1)e^{-\alpha x} + (zB_2 + \alpha A_2)e^{\alpha x} = 0 \\ (yA_1 - \alpha B_1)e^{-\alpha x} + (yA_2 + \alpha B_2)e^{\alpha x} = 0 \end{aligned}\right\}, \tag{51}$$

or noting (43*a*) and (50) we have

$$\left.\begin{aligned} (Z_0B_1 - A_1)e^{-\alpha x} + (Z_0B_2 + A_2)e^{\alpha x} = 0 \\ (A_1 - Z_0B_1)e^{-\alpha x} + (A_2 + Z_0B_2)e^{\alpha x} = 0 \end{aligned}\right\}, \tag{52}$$

where the symbol

$$Z_0 = \sqrt{\frac{z}{y}} = \sqrt{\frac{R + jL\omega}{G + jC\omega}} \tag{53}$$

has been introduced. Since these relations are to hold for all values of x, the coefficients of $e^{-\alpha x}$ and $e^{\alpha x}$ must vanish separately. Making use of this fact we see that the equations (52) are consistent in yielding

$$\left.\begin{aligned} B_1 &= \frac{A_1}{Z_0} \\ B_2 &= \frac{A_2}{-Z_0} \end{aligned}\right\}, \tag{54}$$

so that we now have

$$\left.\begin{aligned} E(x) &= A_1e^{-\alpha x} + A_2e^{\alpha x} \\ I(x) &= \frac{A_1}{Z_0}e^{-\alpha x} - \frac{A_2}{Z_0}e^{\alpha x} \end{aligned}\right\} \tag{55}$$

for the voltage and current amplitudes. These contain only two arbitrary constants A_1 and A_2, which may now be evaluated from boundary conditions in the following manner.

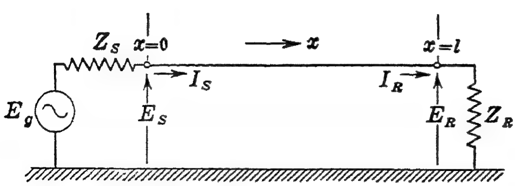


FIG. 4.—Single-line diagram of the transmission line showing terminal conditions and assumed reference directions.

In Fig. 4 is shown the layout of the line with terminal impedances and source of excitation. For this figure the so-called single-line diagram is used which replaces the return conductor by a ground plane. This is merely

for the purpose of simplifying the appearance of the figure. The sending end of the line, where the generator is located, is denoted by $x = 0$; the receiving end is denoted by $x = l$; i.e., distance is measured positively from sending to receiving end, and the line length is denoted by l. Z_S and Z_R are the sending- and receiving-end impedances, respectively, for the angular frequency ω. The positive assumed directions for current are from left to right, and the positive directions for voltage are from the ground plane to the line.

On this basis we may write the following equations at the sending and receiving ends, respectively. The subscripts S and R refer to these ends.

$$\left.\begin{aligned} E_S + I_S Z_S &= E_g \\ E_R - I_R Z_R &= 0 \end{aligned}\right\} \tag{56}$$

By (55) we see that

$$\left.\begin{aligned} E_S &= A_1 + A_2 \\ E_R &= A_1 e^{-\alpha l} + A_2 e^{\alpha l} \\ I_S &= \frac{1}{Z_0}(A_1 - A_2) \\ I_R &= \frac{1}{Z_0}(A_1 e^{-\alpha l} - A_2 e^{\alpha l}) \end{aligned}\right\}. \tag{57}$$

Substituting these into (56), the boundary conditions take the form

$$\left.\begin{aligned} (Z_0 + Z_S)A_1 \qquad + (Z_0 - Z_S)A_2 &= Z_0 E_g \\ (Z_0 - Z_R)e^{-\alpha l}A_1 + (Z_0 + Z_R)e^{\alpha l}A_2 &= 0 \end{aligned}\right\}, \tag{56a}$$

by means of which A_1 and A_2 may be evaluated.

Denoting the determinant of this system of simultaneous equations by D, we have

$$D = (Z_0 + Z_S)(Z_0 + Z_R)e^{\alpha l} - (Z_0 - Z_S)(Z_0 - Z_R)e^{-\alpha l}, \tag{58}$$

and hence

$$\left.\begin{aligned} A_1 &= \frac{(Z_R + Z_0)Z_0 e^{\alpha l} E_g}{D} \\ A_2 &= \frac{(Z_R - Z_0)Z_0 e^{-\alpha l} E_g}{D} \end{aligned}\right\}. \tag{59}$$

Substituting these back into (55), the voltage and current amplitudes become

$$E(x) = \frac{Z_0 E_g[(Z_R + Z_0)e^{\alpha(l-x)} + (Z_R - Z_0)e^{-\alpha(l-x)}]}{(Z_S + Z_0)(Z_R + Z_0)e^{\alpha l} - (Z_S - Z_0)(Z_R - Z_0)e^{-\alpha l}}, \tag{60}$$

and

$$I(x) = \frac{E_g[(Z_R + Z_0)e^{\alpha(l-x)} - (Z_R - Z_0)e^{-\alpha(l-x)}]}{(Z_S + Z_0)(Z_R + Z_0)e^{\alpha l} - (Z_S - Z_0)(Z_R - Z_0)e^{-\alpha l}}, \tag{61}$$

which together with (48), (43a), and (53) constitute the complete steady-state solution desired.

For purposes of interpretation it is convenient to modify the form of the results (60) and (61) in the following way. Introducing the symbols

$$r_S = \frac{Z_S - Z_0}{Z_S + Z_0}, \tag{62}$$

and

$$r_R = \frac{Z_R - Z_0}{Z_R + Z_0}, \tag{63}$$

the expressions (60) and (61) may be put in the form

$$E(x) = \frac{Z_0 E_g (e^{\alpha(l-x)} + r_R e^{-\alpha(l-x)})}{(Z_S + Z_0)(e^{\alpha l} - r_S r_R e^{-\alpha l})}, \tag{60a}$$

and

$$I(x) = \frac{E_g (e^{\alpha(l-x)} - r_R e^{-\alpha(l-x)})}{(Z_S + Z_0)(e^{\alpha l} - r_S r_R e^{-\alpha l})}. \tag{61a}$$

The quantities r_S and r_R may be given a useful physical interpretation. This we wish to take up in more detail now.

5. Wave character of the solution. In this section we wish to show that the steady-state solution found above may be interpreted as physically representing a pair of waves traveling in opposite directions with the same constant velocity. In order to do this it is convenient to make use of the following further notation. Let

$$E^+ = \frac{Z_0 E_g}{(Z_S + Z_0)(e^{\alpha l} - r_S r_R e^{-\alpha l})} \tag{64}$$

and

$$E^- = r_R E^+, \tag{64a}$$

also

$$I^+ = \frac{E^+}{Z_0} \tag{65}$$

and

$$I^- = \frac{E^-}{-Z_0} = -r_R I^+. \tag{65a}$$

Then (60a) and (61a) become

$$\left.\begin{aligned} E(x) &= E^+ e^{\alpha(l-x)} + E^- e^{-\alpha(l-x)} \\ I(x) &= I^+ e^{\alpha(l-x)} + I^- e^{-\alpha(l-x)} \end{aligned}\right\}. \tag{66}$$

Now it is necessary to recognize that the quantity α, defined by (43a), is complex, and hence may be written

$$\alpha = \alpha_1 + j\alpha_2. \tag{43b}$$

α is called the **propagation function** of the line, because, as we shall see, it governs the propagation properties of the system. The real part of α, namely α_1, governs the attenuation of the waves, and hence is called the **attenuation function,** while α_2 will be seen to determine the variation of phases with distance and hence is given the name **phase function.**

One more set of definitions is necessary before the wave expressions may be set up. This involves the complex character of the quantities (64), (64a), (65), and (65a). Here we shall write

$$\left.\begin{aligned} E^+ &= |E^+| \, \underline{/\psi^+} \\ E^- &= |E^-| \, \underline{/\psi^-} \\ I^+ &= |I^+| \, \underline{/\varphi^+} \\ I^- &= |I^-| \, \underline{/\varphi^-} \end{aligned}\right\}. \tag{67}$$

With these notations, the amplitudes (66) may now be substituted into (48) and the real parts evaluated. We find

$$\begin{aligned} e(x,t) = {} & |E^+|e^{\alpha_1(l-x)} \cdot \cos\,[\alpha_2(l-x) + (\omega t + \psi^+)] \\ & + |E^-|e^{-\alpha_1(l-x)} \cdot \cos[\alpha_2(l-x) - (\omega t + \psi^-)], \end{aligned} \tag{68}$$

and

$$\begin{aligned} i(x,t) = {} & |I^+|e^{\alpha_1(l-x)} \cdot \cos\,[\alpha_2(l-x) + (\omega t + \varphi^+)] \\ & + |I^-|e^{-\alpha_1(l-x)} \cdot \cos\,[\alpha_2(l-x) - (\omega t + \varphi^-)]. \end{aligned} \tag{69}$$

The first terms in these expressions represent cosine curves travelling in the positive direction with amplitudes which decay exponentially in the direction of motion. These terms are called the *incident waves.* The second terms are cosine curves traveling in the opposite direction with the same velocity and decaying exponentially in that direction. They are called the *reflected waves,* it being supposed that their presence is due to the fact that a portion of the incident wave is reflected at the receiving end. Noting (64a) and (65a), we see that the magnitudes of the reflected waves at the receiving end (point of reflection) equal the magnitudes of the incident waves at that point multiplied by the magnitude of the factor r_R which is, therefore, called the **reflection coefficient** of the receiving end. The factor r_S, which is called the reflection coefficient of the sending end, does not play the same rôle at that end because the source is assumed located there. If we had considered generators at either end, we would have found a similar reflection phenomenon taking place at the sending end also. The student may set up the boundary conditions for this condition and evaluate the result as an exercise.

The reflection coefficients given by (62) and (63) evidently depend upon the discrepancy between the respective terminal impedance and

the quantity Z_0 defined by (53), which, incidentally, also has the dimensions of an impedance. The latter is characteristic of the line itself and hence is given the name **characteristic impedance.** A physical interpretation of this quantity alone will be given later when a fuller appreciation of the forms (62) and (63) for the reflection coefficients will be attained.

In order to see more clearly that the terms in (68) and (69) referred to above really are traveling cosine curves, the reader may first consider the following simpler expressions:

$$\left.\begin{array}{l}\cos(\alpha_2 x - \omega t) \\ \text{and} \\ \cos(\alpha_2 x + \omega t)\end{array}\right\} \qquad (70)$$

which represent the essential features of the incident and reflected waves respectively, the phases $\alpha_2 l$ and ψ^+, etc., having nothing to do with the general characteristics of these terms. The student should plot the terms (70) as functions of x for successive increasing values of t, obtaining sets of curves for each term. Thus it will become clear that the first progresses in the positive, and the second in the negative, direction. He will then appreciate more fully the wave character of expressions (68) and (69). He should also convince himself of the fact that each component wave (incident and reflected) damps exponentially in the direction of its travel, and that the point of reflection is at the receiving end. There the amplitude of the reflected wave is fixed by that of the incident wave and the reflection coefficient. Unless this point is clearly appreciated, it might be thought that, since the reflected wave grows exponentially in the positive x-direction, its amplitude may increase beyond all bounds.

The only reason for putting the steady-state solution into this form is to aid in an attempt to visualize the resulting phenomenon. Whether or not waves are really present, or what the waves consist of, is altogether immaterial. The wave interpretation should be regarded as a mathematical picture rather than as a physical fact. After all, we can say with certainty only that the results, so far as measurable voltage and current are concerned, *are as though* waves, such as we have described, are present on the system. For this we have mathematical proof, namely in the results (68) and (69). The harmonic steady state is, therefore, as though it were due to the continuous passage of waves in opposite directions at all points in the line, the net voltage or current at any point being given by the linear superposition of the separate contributions of the incident and reflected wave components respectively, where proper attention is paid to the time and space phases of

these components. Thus, in the complex voltage and current amplitudes as expressed by (66), the first terms represent the complex amplitudes of the incident waves at any point x, and the second terms represent the complex amplitudes of the reflected waves. The *vector sum* of the complex amplitudes of the incident and reflected waves at any point gives the net complex amplitude E or I at that point.

These relations (66) also make it clear that the point of reflection is at the receiving end, i.e., that the component amplitudes E^+ and E^- as well as I^+ and I^- refer to that point. To see this, let $x = l$ (which fixes our attention upon the receiving end). Then (66) becomes

$$\left.\begin{aligned} E(l) &= E_R = E^+ + E^- \\ I(l) &= I_R \ = I^+ + I^- \end{aligned}\right\}. \tag{66a}$$

At the sending end, on the other hand, $x = 0$, and we have

$$\left.\begin{aligned} E(0) &= E_S = E^+e^{\alpha l} + E^-e^{-\alpha l} \\ I(0) &= I_S \ = I^+e^{\alpha l} + I^-e^{-\alpha l} \end{aligned}\right\}, \tag{66b}$$

which makes it clear that the complex amplitudes of the incident waves at the sending end are $E^+e^{\alpha l}$ and $I^+e^{\alpha l}$, while those for the reflected waves at that end are $E^-e^{-\alpha l}$ and $I^-e^{-\alpha l}$. This also makes it clear that the ratio of an incident wave amplitude at the sending end to that wave amplitude at the receiving end is $e^{\alpha l}$, and that the corresponding ratio for the reflected wave amplitudes is $e^{-\alpha l}$. The factor $e^{\alpha l}$ or $e^{-\alpha l}$, therefore, represents the change which a component wave amplitude undergoes (both as to magnitude and phase on account of the complex character of α) as it traverses the length of the line. The propagation properties of the line, as they affect the individual wave components, are bound up—so to speak—in the factor $e^{\alpha l}$, or, we may say, in α, which justly, therefore, deserves the name of propagation function as already mentioned.

Another important point which the student should not fail to appreciate in this connection is brought out by the relations (65) and (65a), which show that the ratio of complex voltage and current amplitudes equals $+Z_0$ for the incident waves, and $-Z_0$ for the reflected waves; i.e., the *component* voltage and current wave amplitudes are simply related by the characteristic impedance of the line in the Ohm's law sense, with the novelty that a negative relationship appears with the reflected waves. This simple relationship between voltage and current, however, exists only for the component waves, and the student should not make the mistake of imagining that it, therefore, holds for the net voltage and current at any point! That will be true only under special conditions which cause the reflected wave to be absent. Then the net values equal the incident values so that the net voltage

and current at any point are simply related by the characteristic impedance. This important condition we wish now to discuss in more detail.

6. Characteristic impedance. We mentioned earlier that we would give the reader a more concrete physical notion of how the characteristic impedance (53) may be interpreted. To do this let us consider the expressions (60a) and (61a) for the complex voltage and current amplitudes. In these expressions let us fix our attention particularly upon those terms in which the factor $e^{-\alpha l}$ appears. Recognizing the complex character of α, this factor becomes

$$e^{-\alpha_1 l} \cdot e^{-j\alpha_2 l}.$$

Now suppose that the attenuation function α_1 has some finite value and that the length of the line becomes very large. The factor $e^{-\alpha_1 l}$ then becomes very small. If the line were infinitely long, this factor would vanish altogether, and hence the terms with which it is associated in (60a) and (61a) would also vanish. These are the terms in which r_R appears. When this happens, the resulting expressions for $E(x)$ and $I(x)$ become independent of the line length l because the factors $e^{\alpha l}$ cancel out. This cancelation is perfectly admissible even though l is tending toward infinity. Hence we can say

$$\underset{l\to\infty}{\text{limit}}\left[E(x)\right] = \frac{Z_0 E_g}{Z_S + Z_0} \cdot e^{-\alpha x}, \tag{71}$$

and

$$\underset{l\to\infty}{\text{limit}}\left[I(x)\right] = \frac{E_g}{Z_S + Z_0} \cdot e^{-\alpha x}. \tag{72}$$

Another way of showing this is to factor $e^{\alpha l}$ out of the numerator and denominator of the expressions (60a) and (61a) so that these become

$$E(x) = \frac{Z_0 E_g e^{-\alpha x}}{Z_S + Z_0}\left\{\frac{1 + r_R e^{-2\alpha(l-x)}}{1 - r_S r_R e^{-2\alpha l}}\right\}, \tag{60b}$$

$$I(x) = \frac{E_g e^{-\alpha x}}{Z_S + Z_0}\left\{\frac{1 - r_R e^{-2\alpha(l-x)}}{1 - r_S r_R e^{-2\alpha l}}\right\}. \tag{61b}$$

For large values of l the terms involving $e^{-2\alpha l}$ become small, so that we may expand the bracket terms and neglect all but the linear terms in $e^{-2\alpha l}$. This gives

$$E(x) = \frac{Z_0 E_g e^{-\alpha x}}{Z_S + Z_0}\left\{1 + r_R(r_S + e^{2\alpha x})e^{-2\alpha l} + \cdots\right\}, \tag{60c}$$

$$I(x) = \frac{E_g e^{-\alpha x}}{Z_S + Z_0}\left\{1 + r_R(r_S - e^{2\alpha x})e^{-2\alpha l} + \cdots\right\}. \tag{61c}$$

In the limit $l \to \infty$ all terms in the bracketed series vanish for finite

values of x except the first which are unity, and thus the forms (71) and (72) are obtained. Using these in (48) we see that for an infinitely long line

$$\left.\begin{aligned} e(x,t) &= \Re e\left[\frac{Z_0E_g}{Z_S + Z_0} \cdot e^{-\alpha x + j\omega t}\right] \\ i(x,t) &= \Re e\left[\frac{E_g}{Z_S + Z_0} \cdot e^{-\alpha x + j\omega t}\right] \end{aligned}\right\}, \tag{73}$$

or if we let

$$\frac{Z_0E_g}{Z_S + Z_0} = \left|\frac{Z_0E_g}{Z_S + Z_0}\right| \underline{|\psi'}, \tag{74}$$

and

$$\frac{E_g}{Z_S + Z_0} = \left|\frac{E_g}{Z_S + Z_0}\right| \underline{|\varphi'}, \tag{75}$$

then

$$\left.\begin{aligned} e(x,t) &= \left|\frac{Z_0E_g}{Z_S + Z_0}\right| \cdot e^{-\alpha_1 x} \cdot \cos\left[\alpha_2 x - \omega t - \psi'\right] \\ i(x,t) &= \left|\frac{E_g}{Z_S + Z_0}\right| \cdot e^{-\alpha_1 x} \cdot \cos\left[\alpha_2 x - \omega t - \varphi'\right] \end{aligned}\right\}. \tag{73a}$$

These represent waves traveling in the positive direction. The net voltage and current are, therefore, represented by their incident wave components alone. The reflected waves are absent, as might have been predicted since the point of reflection is infinitely far away. In complete agreement with this situation, we see from (71) and (72) that for this case

$$\frac{E(x)}{I(x)} = Z_0, \tag{76}$$

i.e., the impedance looking into the infinitely long line at any point is equal to Z_0! This gives us a very convenient physical interpretation for the characteristic impedance.

With this in mind, the significance of the reflection coefficients (62) and (63) becomes somewhat clearer also. For example, it is obviously immaterial whether the line is infinitely long or whether it is finite and terminated in a lumped impedance Z_R which is equal to the characteristic impedance, because we have just shown that the latter is equivalent to an infinitely long section of line. Consistent with this idea, we see by (62) that letting $Z_R = Z_0$ makes $r_R = 0$, i.e., there is no reflection at the receiving end under these conditions. If we refer to the expressions (60a) and (61a) for the net voltage and current amplitudes, we see that setting $r_R = 0$ reduces these to the forms (71)

and (72) which were obtained for the infinitely long line. Hence we again see that whether we consider the line infinitely long, or finite but terminated in Z_0, the results are the same. The voltage and current expressions reduce to a single exponential term in either case, and are simply related by the characteristic impedance. In practice this condition is highly desirable for reasons which will be discussed in detail later.

The forms (60*c*) and (61*c*) for the voltage and current amplitudes involving series expansions in terms of the factor $e^{-2\alpha l}$ are interesting from another standpoint. If the line is very long or the attenuation high, the series will converge very rapidly for relatively small values of x. In such cases, therefore, the reflected wave may be neglected to within an approximate error of

$$|r_R(r_S \pm e^{2\alpha x})e^{-2\alpha l}| \cdot 100\%. \tag{77}$$

This is smallest at the sending end, namely

$$|r_R(r_S \pm 1)e^{-2\alpha l}| \cdot 100\%. \tag{77a}$$

For the impedance looking into the line at the sending end we have

$$\frac{E(0)}{I(0)} = Z_0(1 + 2r_R e^{-2\alpha l} + \cdots), \tag{76a}$$

i.e., the impedance looking into the sending end is equal to the characteristic impedance to within an approximate error of

$$2r_R e^{-2\alpha l}\ .$$

This means physically that, under conditions where these approximations are justified, the sending-end impedance may be assumed equal to Z_0 regardless of how the line is terminated! This is useful in many practical cases.

7. Wave length and phase velocity. The wave interpretation, already discussed to some extent above, brings with it certain additional concepts which it is desirable to point out at this time. For this purpose we shall return to the wave solutions (68) and (69). We have mentioned that these are traveling cosine curves which damp exponentially as they go. A wave of this kind, whether it decays or not, has two distinguishing characteristics in addition to its amplitude and phase relative to some arbitrary reference. One of these is made evident if we imagine the time held constant, i.e., consider the cosine function plotted versus x for a constant value of t. If we neglect the damping for the moment, this is a periodic function, of course. It is periodic in x, i.e., distance. A certain length corresponds to a period of the cosine function. This distance obviously is that length for which $\alpha_2 x$

(the variable part of the argument) increases by 2π radians. If we denote this length by λ, we have

$$\alpha_2\lambda = 2\pi$$

$$\lambda = \frac{2\pi}{\alpha_2}. \tag{78}$$

λ is called the **wave length** or also the **spacial period** because it corresponds to a space-period of the cosine function. This result will become better oriented in the reader's mind if he will associate it with another quantity with which he has an acquaintance of longer standing, namely, the temporal period of the usual harmonic function. This is usually written

$$\tau = \frac{1}{f} = \frac{2\pi}{\omega}, \tag{79}$$

where f is the frequency in cycles per second, and ω the angular frequency. Thus λ is for space what τ is for time. Furthermore, the reader should recognize the parallel rôles played by ω and α_2. The first has the dimensions of *radians per second*, while for the second the dimensions are *radians per mile*—or whatever unit is used for distance. We see that α_2 is for distance what ω is for time! This considerably clarifies the significance of α_2—the phase function. It is equal to the rate at which the phase of a component wave changes with distance at any instant of time.

The analogous quantities ω and α_2 may be still further correlated in the following way. Consider, for example, the voltage wave solution. We mentioned that this is a pair of traveling waves. Suppose we now ask: How fast do these waves travel? This is the same as to say: What is the velocity of a given point on either one of the cosine curves? Fixing our attention upon a given point on one of the cosine curves as it moves is equivalent to observing what happens when the argument of that cosine as a whole remains constant while the variables x and t in it change. This may be expressed mathematically by setting the differential of the argument equal to zero. Doing this for the incident wave, we have

$$d[\alpha_2(l - x) + (\omega t + \psi^+)] = 0,$$

which gives

$$-\alpha_2 dx + \omega dt = 0,$$

or

$$\alpha_2 dx = \omega dt,$$

and thus we have

$$\frac{dx}{dt} = \frac{\omega}{\alpha_2}. \tag{80}$$

This states that, as we fix our attention upon any point on the incident wave, we observe that the rate at which distance changes with time is given by ω/α_2. But this is the velocity with which the entire wave moves, namely,

$$v = \frac{\omega}{\alpha_2} \tag{80a}$$

for the incident wave.

For the reflected wave we form

$$d[\alpha_2(l - x) - (\omega t + \psi^-)] = 0,$$

and get

$$\frac{dx}{dt} = -\frac{\omega}{\alpha_2}, \tag{81}$$

which shows that the reflected wave moves in the negative direction with the same velocity. The velocity with which the component waves move is, therefore, the correlation factor between ω and α_2. We can also see that this result checks dimensionally because *radians per second* divided by *radians per mile* gives *miles per second.*

The reader should bear in mind in this connection that this velocity of propagation of the component waves in the harmonic steady state is not necessarily the velocity at which electromagnetic energy is propagated along the line. We mentioned earlier that the wave concept introduced in this discussion is merely a mathematical fiction which is injected into the picture for convenience in interpreting the results. The student should not make the mistake of becoming so wedded to this mode of interpretation as to forget entirely its fictitious character. If he will stop to reflect for a moment he will realize that questions regarding the propagation of energy, or voltage and current for that matter, have no utility in connection with the steady-state discussion. Velocity of propagation in the usual sense has to do with something that has a beginning and an end, and a steady state has, by definition, neither beginning nor end. Such things as velocities of wave fronts or of wave groups are part of the transient behavior to which we shall come later. In order to avoid confusion on this point, the velocity (80a), which refers only to the fictitious steady-state waves, is referred to as the **phase velocity** since it is the velocity at which a point of constant phase, for either component wave, is propagated.

8. Hyperbolic form. In the discussion so far, the solution was, for the most part, left in the exponential form. This form will usually be found most convenient for numerical calculation. There is another form in terms of hyperbolic functions, however, which finds favor in a large part of the literature on this subject, and it is for the purpose of

acquainting the reader with its relation to the exponential form that the following is given.

The hyperbolic form follows almost immediately from the relations (60) and (61) which were obtained directly after the evaluation of integration constants. If we multiply out the factored terms in the numerator and denominator of each expression, and then factor so as to form the groups

$$\begin{aligned} e^{\alpha(l-x)} + e^{-\alpha(l-x)} &= 2\cosh\alpha(l-x) \\ e^{\alpha(l-x)} - e^{-\alpha(l-x)} &= 2\sinh\alpha(l-x) \\ e^{\alpha l} + e^{-\alpha l} &= 2\cosh\alpha l \\ e^{\alpha l} - e^{-\alpha l} &= 2\sinh\alpha l \end{aligned}$$

we obtain

$$E(x) = \frac{Z_0E_g\{Z_R\cosh\alpha(l-x) + Z_0\sinh\alpha(l-x)\}}{Z_0(Z_S+Z_R)\cosh\alpha l + (Z_SZ_R+Z_0^2)\sinh\alpha l} \tag{60d}$$

$$I(x) = \frac{E_g\{Z_0\cosh\alpha(l-x) + Z_R\sinh\alpha(l-x)\}}{Z_0(Z_S+Z_R)\cosh\alpha l + (Z_SZ_R+Z_0^2)\sinh\alpha l} \tag{61d}$$

which are the desired forms.

When charts for hyperbolic functions of complex arguments are available, these results may be quite convenient for numerical calculation of the complex voltage and current amplitudes. Under ordinary circumstances, however, the exponential forms are of greater utility.

9. Symmetrical forms. In general an almost endless number of variations may be given to the form of the steady-state solution, as the reader will find if he pursues the literature on this subject somewhat. A certain amount of personal taste enters into the situation, and no two authors will be found to agree exactly as to form or notation. Obviously nothing much is to be gained by giving additional variations, except to point out one which adds much to the compactness of the resulting expressions, and may save time in numerical calculations.

The reader will probably have noticed one obvious defect in the forms given so far, and that is that they lack symmetry. The complex forms (60a) and (61a), for example, are marred by the appearance of the factors r_R and r_Sr_R in connection with one of the exponentials in each pair of terms of the numerator and denominator. But for this lack of symmetry, these pairs could be contracted into a single hyperbolic function. This lack of symmetry may be removed by the following simple expedient.

If we factor $r_R^{1/2}$ out of the numerator and $r_S^{1/2}r_R^{1/2}$ out of the denominator, we will have for the voltage amplitude

$$E(x) = \frac{Z_0E_g(r_R^{-1/2}e^{\alpha(l-x)} + r_R^{1/2}e^{-\alpha(l-x)})}{r_S^{1/2}(Z_S+Z_0)(r_S^{-1/2}r_R^{-1/2}e^{\alpha l} - r_S^{1/2}r_R^{1/2}e^{-\alpha l})}, \tag{82}$$

for which the numerator and denominator now contain symmetrical groups. In order to make this still more evident, we define the quantities ρ_S and ρ_R such that

$$\left.\begin{aligned} r_S^{1/2} &= e^{\rho_S};\ \rho_S = \tfrac{1}{2}\ln r_S \\ r_R^{1/2} &= e^{\rho_R};\ \rho_R = \tfrac{1}{2}\ln r_R \end{aligned}\right\}. \tag{83}$$

Then the numerator and denominator consist of symmetrical exponential groups for which we can substitute the hyperbolic functions and obtain

$$E(x) = \frac{Z_0 E_g \cosh\left[\alpha(l - x) - \rho_R\right]}{\sqrt{Z_S^2 - Z_0^2}\sinh\left[\alpha l - \rho_S - \rho_R\right]}. \tag{60e}$$

Similarly the current amplitude becomes

$$I(x) = \frac{E_g \sinh\left[\alpha(l - x) - \rho_R\right]}{\sqrt{Z_S^2 - Z_0^2}\sinh\left[\alpha l - \rho_S - \rho_R\right]}. \tag{61e}$$

In both of these the expression (62) for r_S has also been made use of. Thus the effect of the terminal impedances is taken care of by the quantities ρ_S and ρ_R in the arguments of the hyperbolic functions. These quantities are sometimes referred to as the *equivalent line angles*, it being imagined that the terminal impedances have been replaced by fictitious extensions of the line. The resulting expressions are extremely compact. If hyperbolic charts are available, these forms may save time in computation, although the equivalent line angles must first be determined from (83).

10. Neglected dissipation. In the next four sections we shall discuss primarily the general characteristics of the steady-state solution for a line whose resistance and leakage are negligible. It may be argued that such a case does not occur in practice and that, therefore, its discussion is a waste of time.[1] The reader should bear in mind, however, that the performance of an actual line is best understood in the light of a modification of the behavior of the corresponding dissipationless line. In other words, the effect of resistance and leakage can be appreciated only after some familiarity has been gained with a system in which these parameters are absent. Only in the dissipationless case do some of the most interesting features concerning the steady-state line performance appear in their true light. The introduction of the parameters R and G, if sufficient in magnitude, may entirely obscure some

[1] In this connection it should be pointed out that high-frequency transmission lines used for feeding antennae are quite commonly treated by neglecting dissipation. See, for example, Hans Roder, "Graphical Methods for Problems Involving Radio-Frequency Transmission Lines," I.R.E. Proc., Vol. 21, No. 2, Feb., 1933, p. 290.

of the line's most peculiar properties which the student should not fail to appreciate. The effect of these parameters will, therefore, be taken up subsequently.

One of the first points of interest in a line without loss concerns itself with the propagation function (43a). Here we see that for $R = G = 0$ we get

$$\alpha = j\omega\sqrt{LC}, \tag{43b}$$

so that

$$\left.\begin{aligned} \alpha_1 &= 0 \\ \alpha_2 &= \omega\sqrt{LC} \end{aligned}\right\}. \tag{43c}$$

The component waves are, therefore, not attenuated as they travel. By (80a) the velocity of these waves is

$$v = \frac{\omega}{\alpha_2} = \frac{1}{\sqrt{LC}}, \tag{80b}$$

which, according to the discussion given in the first chapter, equals the velocity of light for open lines. The important point, however, is that this velocity is independent of frequency. This means that in the dissipationless case all frequencies have the same phase velocity.

The characteristic impedance according to (53) also becomes particularly simple for $R = G = 0$, namely

$$Z_0 = \sqrt{\frac{L}{C}}, \tag{53a}$$

which is a constant independent of frequency and hence is in effect a pure resistance. Thus for this case it is possible to terminate the line in its characteristic impedance. If this were done we would have by (74), (75), and (73a), assuming a pure resistance for Z_S and taking E_g as a reference

$$\left.\begin{aligned} e(x, t) &= \frac{Z_0E_g}{Z_S + Z_0}\cos\omega(t - x\sqrt{LC}) \\ i(x, t) &= \frac{E_g}{Z_S + Z_0}\cos\omega(t - x\sqrt{LC}) \end{aligned}\right\}, \tag{73b}$$

where (43c) has been taken into account. This result makes it clear that, when a dissipationless line is properly terminated, it becomes a distortionless transmission system for the reason that neither the amplitude nor phase of the voltage or current is a function of frequency! This means that, when a complex periodic voltage is impressed, all its harmonic components will receive the same amplitude and phase changes during the process of transmission, so that the periodic function

at the receiving end will be an exact replica of that which was impressed. Later we shall see that transient functions may also be looked upon as the result of an infinite number of steady harmonic functions linearly superposed. Hence this distortionless property of the dissipationless line (when properly terminated) applies to all classes of impressed forces which may occur in practice.

The student should note well that the distortionless property of the dissipationless line does not follow unless the line is properly terminated, i.e., terminated in its characteristic impedance (53a). In order to convince himself of this he should introduce the relations (43c) into the general expressions (60a) and (61a), or their equivalents, and then form the instantaneous voltage and current expressions. He will find that unless $r_R = 0$, the resulting system will still possess distortion even though the propagation function is ideal. In general a line is used for communication in both directions, so that it will be necessary to properly terminate at both ends, since the sending and receiving ends continually interchange rôles. Proper termination is also essential from the standpoint of suppressing reflections which might otherwise cause echoes as well as interference with neighboring circuits.

It is also interesting to study the current and voltage relationships on the dissipationless line. Since these, however, involve the terminal impedances, we must first decide upon the nature of the latter before we can proceed with a more detailed discussion. Here two cases are of special interest. One is for pure reactive terminal impedances, i.e., for the case where the entire system, inclusive of its terminations, is dissipationless. The other is for pure resistance terminations associated with the dissipationless line. These we shall take up in the order named.

When Z_S and Z_R are pure reactances they are pure imaginary functions of frequency. In order to simplify the equations, let us write

$$\left.\begin{aligned} \frac{Z_S}{Z_0} &= jk_S \\ \frac{Z_R}{Z_0} &= jk_R \end{aligned}\right\}. \tag{84}$$

Then we have for the reflection coefficients (62) and (63)

$$r_S = \frac{jk_S - 1}{jk_S + 1}, \tag{62a}$$

and

$$r_R = \frac{jk_R - 1}{jk_R + 1}. \tag{63a}$$

But these are complex numbers whose magnitudes are unity for all

values of k_S or k_R, and hence for all frequencies. The reflection coefficients are, therefore, pure rotational operators. If in this case we let

$$\left.\begin{aligned} r_S^{1/2} &= e^{j\vartheta_S} \\ r_R^{1/2} &= e^{j\vartheta_R} \end{aligned}\right\}, \tag{85}$$

then

$$\left.\begin{aligned} \vartheta_S &= \cot^{-1} k_S \\ \vartheta_R &= \cot^{-1} k_R \end{aligned}\right\}, \tag{85a}$$

and from (83) we see that

$$\begin{aligned} \rho_S &= j\vartheta_S \\ \rho_R &= j\vartheta_R \end{aligned} \tag{83a}$$

so that the forms (60e) and (61e) for the complex voltage and current amplitudes become

$$E(x) = \frac{-E_g \cos\,[\alpha_2(l - x) - \vartheta_R]}{\sqrt{1 + k_S^2}\,\sin\,(\alpha_2 l - \vartheta_S - \vartheta_R)}, \tag{60f}$$

and

$$I(x) = \frac{-jE_g \sin\,[\alpha_2(l - x) - \vartheta_R]}{Z_0\sqrt{1 + k_S^2}\,\sin\,(\alpha_2 l - \vartheta_S - \vartheta_R)}. \tag{61f}$$

For the exponential forms we see that by treating (64) in a similar manner to (82) we have

$$E^+ = \frac{-E_g e^{-j\vartheta_R}}{2\sqrt{1 + k_S^2}\,\sin\,(\alpha_2 l - \vartheta_S - \vartheta_R)}, \tag{64b}$$

and

$$E^- = E^+ e^{2j\vartheta_R}. \tag{64c}$$

The incident and reflected wave amplitudes for the current are these divided by $\pm Z_0$ respectively. Note that, if E_g is taken as a reference vector, E^+ and E^- form a conjugate pair, while I^- is the negative conjugate of I^+. The exponential forms (66) become

$$\left.\begin{aligned} E(x) &= E^+ e^{j\alpha_2(l-x)} + E^- e^{-j\alpha_2(l-x)} \\ I(x) &= I^+ e^{j\alpha_2(l-x)} + I^- e^{-j\alpha_2(l-x)} \end{aligned}\right\}. \tag{66c}$$

Since these are in the form of the sum and difference of conjugate pairs respectively, their values equal 2 times the real part and $2j$ times the imaginary part of the first terms. The student may carry this out as an exercise and thus verify the results (60*f*) and (61*f*). He should note that in this case, where the entire system is dissipationless, the complex voltage and current amplitudes (60*f*) and (61*f*) are in quadrature at all points along the line, and also that, when E_g is taken

as a reference, E is purely real while I is a pure imaginary. Thus the power product is zero at all points in the system, as should be expected.

When the line itself is dissipationless, but terminated in pure resistances, the reflection coefficients (62) and (63) are real numbers whose values lie between plus and minus one, the first occurring for an open end and the second for a short-circuited end. The propagation function is, of course, pure imaginary. With these things in mind, the student can easily write down the expressions for this case using any of the above forms for the general one.

11. Standing waves. An interesting phenomenon occurs in the complete dissipationless system discussed in the previous section. Using (60f) and (61f) to form the instantaneous voltage and current according to (48), we have

$$e(x, t) = \frac{-E_g \cos [\alpha_2(l - x) - \vartheta_R] \cdot \cos \omega t}{\sqrt{1 + k_S^2} \sin (\alpha_2 l - \vartheta_S - \vartheta_R)}, \tag{68a}$$

and

$$i(x, t) = \frac{E_g \sin [\alpha_2(l - x) - \vartheta_R] \cdot \sin \omega t}{Z_0\sqrt{1 + k_S^2} \sin (\alpha_2 l - \vartheta_S - \vartheta_R)}. \tag{69a}$$

These are harmonic functions of time whose amplitudes are harmonic functions of distance along the line. The voltage and current functions

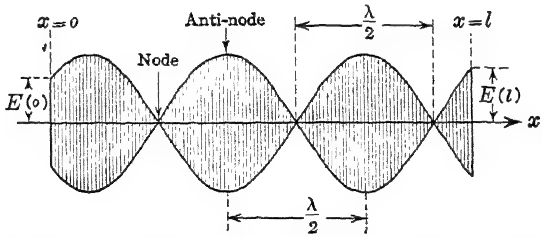


FIG. 5.—Diagrammatic representation of a standing wave.

are both in space and time quadrature. The time phase of each function is the same everywhere along the line. Thus the harmonic time-variation is in unison throughout the line, and its amplitude varies harmonically along the line. There are points in the line where the amplitude is a maximum, and other points where the amplitude is zero at all times. The amplitude variation involved in this phenomenon is illustrated in Fig. 5. The harmonic envelope indicates the limits between which the synchronous harmonic time variation takes place. The result gives the appearance of a harmonic wave which is fixed in position, i.e., does not move. The phenomenon is therefore referred to as a **standing wave.** The point at which the amplitude is a maximum

is called an **anti-node**, while the zero points are called **nodes**. The distance between any pair of nodes or anti-nodes is obviously one-half a wave length. The phenomenon of standing waves thus places the significance of the wave length in evidence. This fact, together with the relation (78), which for the dissipationless case becomes

$$\lambda = \frac{2\pi}{\omega\sqrt{LC}} = \frac{\tau}{\sqrt{LC}} = \frac{v}{f}, \tag{78a}$$

is made use of in the laboratory for the measurement of high frequencies. A pair of large, well-insulated conductors are set up and coupled to a source of unknown frequency. A Geissler tube, whose length equals the spacing between conductors, is held so as to bridge these, and moved slowly along the line. By noting the points at which the tube extinguishes, the distance between two successive nodes is measured, thus determining λ. Assuming the velocity of light for v, the frequency can then be calculated. Such a system of parallel wires is called a *Lecher wire system*.[1]

Note that the amplitude of an anti-node is determined primarily by the factor

$$\sin(\alpha_2 l - \vartheta_S - \vartheta_R)$$

in the denominator of either (68a) or (69a). If the length of the line and the terminal impedances are properly chosen, the argument of this sine function can be made equal to an integer number of π's. The sine will then vanish, and the anti-nodes will theoretically become infinite. In any physical system, however, the incidental dissipation in the line and in the source will keep their magnitudes finite, but they will be a maximum for such an adjustment. In a Lecher wire system the ends are either left open or short-circuited, whichever is most convenient for a particular case, and the length of the line is made adjustable by means of a sliding mechanism of some kind so that the maximum condition just referred to can be obtained by trial. This adjustment obviously gives the maximum sensitivity for the device.

The student should note that when the far end is open the voltage has an anti-node there while the current has a node. With the far end short-circuited, the situation is the reverse. The reader may study various special cases as an exercise.

Before leaving the subject of standing waves, we wish to point out that this result in no way contradicts what has been said previously regarding the general wave character of the solution. The fact that, under special conditions, the net behavior has the appearance of a

[1] Lecher, Wied. Ann. 41, 1890, pp. 850–870.

standing wave does not mean that the incident and reflected wave components no longer move, or that one of them has disappeared and the other become stationary. Both are still there. The fact that the line is dissipationless means that neither wave attenuates. Owing to the non-dissipative terminal networks, the reflection coefficients have unit magnitude so that the phenomenon of reflection causes no change in amplitude but merely a change in phase. The incident and reflected waves, therefore, have the same amplitudes and do not attenuate. The phenomenon of two such unattenuated harmonic waves traveling in opposite directions with the same velocity gives rise to the standing wave effect. The phases of the waves with respect to a given point are of no consequence. The fact that their velocities are the same causes the net effect to be stationary. If one velocity were greater, then the net result would move in that direction.

12. Polar plots. Since $E(x)$ and $I(x)$ are vectors, the loci of their tips as functions of distance along the line for given terminal conditions are given by polar diagrams. In these diagrams the radius vectors represent the magnitudes and phase angles of the voltage and current at various points along the line. They are, therefore, particularly effective in lending pictorial significance to the line behavior.

An idea of the general character of the polar plots can best be obtained from a consideration of the forms (66) which represent the net voltage and current vectors in terms of the vector sums of incident and reflected vector-components. The individual polar plots of these component vectors are obviously logarithmic spirals. For example, we have for the incident voltage vector

$$E^+ \cdot e^{\alpha(l-x)} = E^+ \cdot e^{\alpha_1(l-x)} \cdot e^{j\alpha_2(l-x)}.$$

Starting with $x = 0$ and proceeding toward $x = l$, this represents a vector which decays exponentially and rotates uniformly in the clockwise direction, the rates of decay and rotation being determined by α_1 and α_2 respectively. The tip of this vector traces a logarithmic spiral.

The reflected voltage vector

$$E^- \cdot e^{-\alpha(l-x)} = E^- \cdot e^{-\alpha_1(l-x)} \cdot e^{-j\alpha_2(l-x)}$$

may be represented in a similar manner. This vector, however, decays exponentially and rotates uniformly in a clockwise direction as x is allowed to *decrease* from $x = l$ to $x = 0$. At the end $x = l$ the incident vector is E^+ and the reflected vector is E^-. If the reflection coefficient r_R has a magnitude different from unity and also has an angle, then E^- will be different from E^+ both in magnitude and angle.

The net voltage at any point in the line is the vector sum of the vectors representing the incident and reflected waves at that point.

If the logarithmic spirals for these are superposed, and a sufficient number of points corresponding to arbitrarily chosen x-values marked on each, then the resultant vectors for these points may be found and a continuous curve drawn through their tips. This resultant polar plot will show the variation in magnitude and phase of the actual line voltage with respect to distance. Its shape is that of an elliptic spiral, although it may be difficult to recognize this if the attenuation is high.

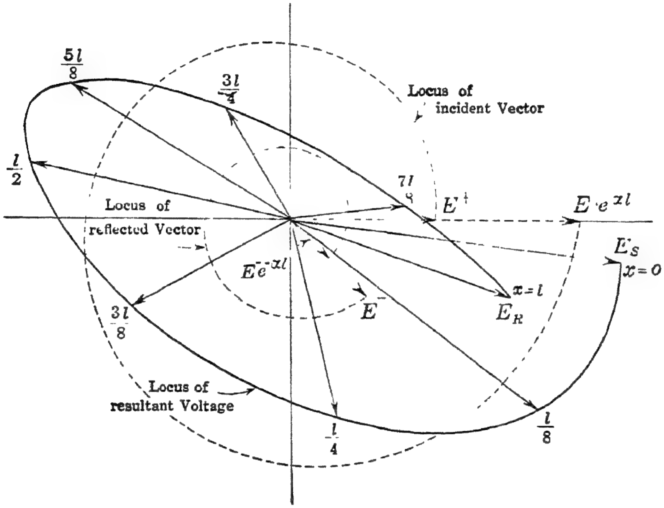


FIG. 6.—Polar plot showing the variation of the vector voltage along a dissipative line one wave length long.

The same procedure may be carried out for the current. Points may be marked on the resultant plots corresponding to fractions of the total line length measured from the sending end. If a sufficient number of points are thus indicated, the plots may be used to read graphically the magnitudes and phase angles of $E(x)$ and $I(x)$ at any desired point in the line. Fig. 6 illustrates the polar plot for voltage for the following arbitrarily chosen data

$$\left.\begin{array}{l} l = \lambda;\ \alpha_2 l = 2\pi \\ \alpha_1 = 0.1104\alpha_2;\ e^{\alpha_1 l} = 2 \\ E^+ \text{ taken as reference} \\ r_R = 0.75 \,\underline{\big|-\frac{\pi}{4}} \end{array}\right\}.$$

Several points may be emphasized with regard to this figure. The student should note well that although the component vectors rotate

uniformly with distance, the resultant does not. In the figure the resultant voltage vectors are drawn for each eighth of a line length. It is quite apparent that the angles between the various vectors are not equal; nor is the total phase change in the net voltage 2π radians although the line is one wave length long. The incident and reflected components, on the other hand, rotate through $\pi/4$ radians for each interval $l/8$. It is also clear from the figure that the reflected vector has its greatest effect in the vicinity of the receiving end of the line. If the figure were carried out for an additional wave length, the resultant locus would very nearly coincide with that of the incident vector in the vicinity of the sending end. When the line is terminated in its characteristic impedance, the net voltage locus becomes the logarithmic spiral representing the incident component.

When line losses are neglected, the situation regarding polar plots becomes very much simpler. The fact that this case is of practical as well as theoretical importance has already been pointed out. When the attenuation on the line is zero, the loci for the incident and reflected vector-components become circles, and the resultant locus is an ellipse. This may be shown in the following way:

Using the relations (66) and (64a) we may write for the voltage in the non-dissipative case

$$E(x) = E^+(e^{j\alpha_2(l-x)} + r_R e^{-j\alpha_2(l-x)}). \tag{86}$$

Assuming an arbitrary terminal impedance Z_R, the reflection coefficient r_R will be complex. If we write

$$r_R = |r_R|\, e^{2j\vartheta_R}, \tag{87}$$

then (86) becomes

$$E(x) = E^+ \cdot e^{j\vartheta_R}(e^{j[\alpha_2(l-x)-\vartheta_R]} + |r_R|\, e^{-j[\alpha_2(l-x)-\vartheta_R]}). \tag{88}$$

Since $E^+ \cdot e^{j\vartheta_R}$ is a factor which is independent of x, it will be more convenient and just as useful to discuss the ratio

$$\frac{E(x)}{E^+} \cdot e^{-j\vartheta_R} = \mathcal{E} = \mathcal{E}_1 + j\mathcal{E}_2 \tag{89}$$

in place of the voltage E itself. The latter is simply rotated and stretched according to the angle and magnitude of $E^+ \cdot e^{j\vartheta_R}$. Since

$$E(l) = E_R = E^+(1 + r_R), \tag{90}$$

the relation (88) may also be expressed in terms of the voltage at the receiving end instead of E^+.

Combining (88) and (89) and writing sines and cosines for the exponentials, we get by equating reals and imaginaries

$$\left.\begin{aligned} \mathcal{E}_1 &= (1 + |r_R|) \cos [\alpha_2(l - x) - \vartheta_R] \\ \mathcal{E}_2 &= (1 - |r_R|) \sin [\alpha_2(l - x) - \vartheta_R] \end{aligned}\right\}. \tag{91}$$

Between these equations it is possible to eliminate either $|r_R|$ or x. In order to do the latter, we divide the first equation by $(1 + |r_R|)$, the second by $(1 - |r_R|)$, square each separately, and then add. We then have

$$\frac{\mathcal{E}_1^2}{(1 + |r_R|)^2} + \frac{\mathcal{E}_2^2}{(1 - |r_R|)^2} = 1, \tag{92}$$

which is the equation of an ellipse in rectangular coordinates with the semi-major axis $(1 + |r_R|)$, and semi-minor axis $(1 - |r_R|)$. Every point on this ellipse corresponds to a terminal impedance giving rise to the same $|r_R|$, but to a definite point in the line. That is, various points on the contour of the ellipse represent various points along the line. We have yet to show how a point on the ellipse may be located to correspond to a given point in the line. This correlation is given by the following.

Suppose we eliminate $|r_R|$ from (91) by dividing by the trigonometric functions and then adding, thus

$$\frac{\mathcal{E}_1}{\cos [\alpha_2(l - x) - \vartheta_R]} + \frac{\mathcal{E}_2}{\sin [\alpha_2(l - x) - \vartheta_R]} = 2. \tag{93}$$

This is the parametric equation of a straight line in cartesian coordinates. It is a straight line with the intercepts $2\cos [\alpha_2(l-x)-\vartheta_R]$ on the horizontal, and $2\sin [\alpha_2(l-x)-\vartheta_R]$ on the vertical axis. The constant distance between these intercepts is 2. The angle $[\alpha_2(l-x)-\vartheta_R]$ is that included between the line and the horizontal axis. This line may be looked upon as a locus on which $|r_R|$ varies from point to point while x and ϑ_R remain constant. Thus if the ellipse (92) and the line (93) are superposed, their intersection determines the vector $\mathcal{E}$ corresponding to the values x, ϑ_R, and $|r_R|$ for which the loci are drawn.

It is well known in geometry, however, that a fixed point on the line (93) will trace an ellipse when the angle of the sine and cosine functions is allowed to pass from zero to 2π, the sum of the major and minor axes of this ellipse being 2. This is none other than the ellipse (92) when the fixed point is so located that its distances from the intercepts with the horizontal and vertical axes are $(1-|r_R|)$ and $(1 + |r_R|)$ respectively. Thus the line (98) may be used to generate the desired locus (92); and the distance x is thus correlated with each point in this

locus by the angle $[\alpha_2(l-x)-\vartheta_R]$. This is illustrated in Fig. 7. The length PQ is constant and equal to 2. The point R which is determined from $|r_R|$ traces the ellipse as x varies, i.e., as the line is rotated so as to keep P and Q on the vertical and horizontal axes respectively.

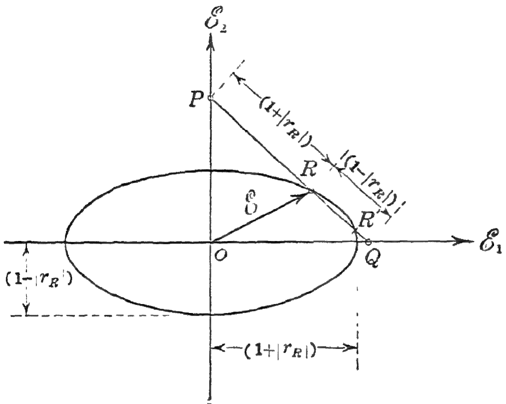


FIG. 7.—Elliptic polar plot of the voltage variation along a non-dissipative line with arbitrary terminal impedance.

This method of correlation makes the geometric construction very simple. In fact, to find $\mathcal{E}$ for a given x and r_R, it is not necessary to draw the ellipse at all. The line PQ, two units in length, is oriented according to the given x and ϑ_R. The point R is located by making PR equal to $(1+|r_R|)$. The vector OR is then the desired $\mathcal{E}$.

When line losses are neglected, Z_0 is real, and hence the largest value which $|r_R|$ can have is unity. The range zero-to-unity for $|r_R|$, therefore, exhausts all the possibilities for the case under discussion. This means that the point R will always lie between P and Q. In the limiting case $|r_R| = 1$, which occurs for either the open or short-circuited line, R coincides with Q, and the ellipse degenerates into a double straight line lying in the horizontal axis. The same occurs for any pure reactive termination. For $Z_R = Z_0$ the reflection coefficient vanishes, and R lies midway between P and Q. The resulting locus is a circle, as is to be expected. Fig. 8 shows a family of loci for various values of terminal impedance. The horizontal axis between mm' is the degenerate ellipse.

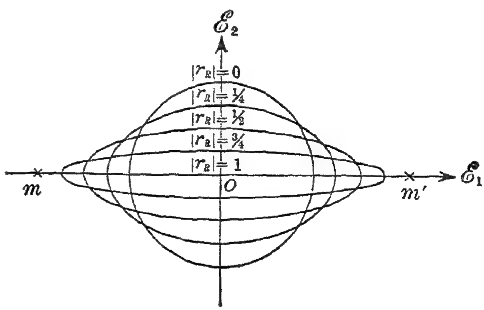


FIG. 8.—Elliptic voltage loci for various values of terminal impedance.

When the line PQ is superposed upon the family of loci of Fig. 8, it may seem confusing to pick the proper intersections with the various

ellipses owing to the fact that the line will have a second intersection with some of these, as for example at the point R' in Fig. 7. Such an intersection must be disregarded on the ground that the point in question does not generate the corresponding ellipse, as may easily be seen.

The loci giving the current variation along the line may be obtained in the same fashion. Using (65), (65*a*), and (66) we have

$$I(x) = \frac{E^+ e^{j\vartheta_R}}{Z_0} \left(e^{j[\alpha_2(l-x)-\vartheta_R]} - |r_R|\, e^{-j[\alpha_2(l-x)-\vartheta_R]}\right) \tag{94}$$

in which (87) is again substituted for r_R. Defining a vector $\mathfrak{J}$ by the relation

$$\frac{Z_0 I}{E^+ e^{j\vartheta_R}} = \mathfrak{J} = \mathfrak{J}_1 + j\mathfrak{J}_2, \tag{95}$$

we have

$$\left.\begin{aligned} \mathfrak{J}_1 &= (1 - |r_R|) \cos\,[\alpha_2(l - x) - \vartheta_R] \\ \mathfrak{J}_2 &= (1 + |r_R|) \sin\,[\alpha_2(l - x) - \vartheta_R] \end{aligned}\right\}, \tag{96}$$

from which we obtain the following equations for the ellipse and straight line

$$\frac{\mathfrak{J}_1^2}{(1 - |r_R|)^2} + \frac{\mathfrak{J}_2^2}{(1 + |r_R|)^2} = 1 \tag{97}$$

$$\frac{\mathfrak{J}_1}{\cos\,[\alpha_2(l - x) - \vartheta_R]} + \frac{\mathfrak{J}_2}{\sin\,[\alpha_2(l - x) - \vartheta_R]} = 2. \tag{98}$$

Comparing these with (92) and (93) we see that the corresponding current and voltage ellipses have their major and minor axes interchanged, while the straight lines are the same for both voltage and current. Whereas for the voltage ellipse the major axis is horizontal, it coincides with the vertical axis for the current ellipse. The method of constructing the current loci needs no additional comment.[1]

These polar plots show that whenever the line is not terminated in its characteristic impedance the voltage and current magnitudes alter-

[1] It is interesting to note that, in the case of real values of Z_R (pure resistance termination), the above relations become particularly simple. We then have for the voltage and current vectors in terms of the receiving-end voltage and current

$$\left.\begin{aligned} E(x) &= E_R \cos \alpha_2(l-x) + jE_R \cdot \frac{Z_0}{Z_R} \sin \alpha_2(l-x) \\ I(x) &= I_R \cos \alpha_2(l-x) + jI_R \cdot \frac{Z_R}{Z_0} \sin \alpha_2(l-x) \end{aligned}\right\}.$$

Taking E_R as a reference, the first of these represents an ellipse with E_R as its semi-axis in the horizontal direction and $E_R \cdot \dfrac{Z_0}{Z_R}$ in the vertical direction. A similar situation holds for the current ellipse with I_R as reference.

nately pass through maximum and minimum values as the length of the line is traversed, the magnitude of the variation being larger for larger departures from proper termination. When line losses are neglected, the maxima and minima occur at quarter-wave-length intervals. The fact that the major axes of the ellipse diagrams for voltage and current are in quadrature means that the respective maxima and minima replace each other; i.e., the voltage has a maximum where the current has a minimum, and vice versa.

This phenomenon of the appearance of maxima and minima, which is accentuated in standing waves, is called the **Ferranti effect.** The presence of dissipation, if sufficient in magnitude, will tend to suppress the effect more or less completely.

13. Odd and even quarter-wave-length lines. When, in the non-dissipative case, the line length equals either an odd or even multiple of a quarter wave length, some peculiar characteristics appear.

Referring to the general expressions (60a) and (61a) for the complex voltage and current amplitudes, we obtain the following expressions for ratios of receiving-to-sending-end quantities for a non-dissipative line with arbitrary terminal impedances

$$\frac{E(l)}{E(0)} = \frac{1 + r_R}{e^{j\alpha_2 l} + r_R e^{-j\alpha_2 l}}, \tag{99}$$

and

$$\frac{I(l)}{E(0)} = \frac{1 - r_R}{Z_0(e^{j\alpha_2 l} + r_R e^{-j\alpha_2 l})}. \tag{100}$$

For

$$l = \frac{\nu\lambda}{4} = \frac{\nu\pi}{2\alpha_2}; \ (\nu = 1, 3, 5, \cdots), \tag{101}$$

we have

$$e^{j\alpha_2 l} + r_R e^{-j\alpha_2 l} = e^{\frac{j\nu\pi}{2}}(1 + r_R e^{-j\nu\pi}) = j(1 - r_R)(-1)^{\frac{\nu-1}{2}}. \tag{102}$$

Noting that by (63)

$$\frac{1 + r_R}{1 - r_R} = \frac{Z_R}{Z_0}, \tag{103}$$

we get for this case

$$\frac{E(l)}{E(0)} = j\frac{Z_R}{Z_0} \cdot (-1)^{\frac{\nu+1}{2}}, \tag{99a}$$

and

$$\frac{I(l)}{E(0)} = j\frac{(-1)^{\frac{\nu+1}{2}}}{Z_0}. \tag{100a}$$

On the other hand for

$$l = \frac{\nu\lambda}{4} = \frac{\nu\pi}{2\alpha_2}; \; (\nu = 0, 2, 4, \cdots), \tag{101a}$$

we have

$$e^{j\alpha_2 l} + r_R e^{-j\alpha_2 l} = (1 + r_R)(-1)^{\frac{\nu}{2}}, \tag{102a}$$

and hence

$$\frac{E(l)}{E(0)} = (-1)^{\frac{\nu+2}{2}}, \tag{99b}$$

and

$$\frac{I(l)}{E(0)} = \frac{Z_0}{Z_R}(-1)^{\frac{\nu+2}{2}}. \tag{100b}$$

For the line lengths (101), for which (99*a*) and (100*a*) follow, we see that if we consider the sending-end voltage $E(0)$ constant, the received voltage becomes a linear function of the load impedance Z_R, while the load current becomes independent of the load impedance, i.e., a constant! Thus the line will virtually transform a constant-voltage system into a constant-current system. For the line lengths (101*a*), on the other hand, the received voltage becomes independent of the load impedance, while the current varies inversely as that impedance. Here the effect of the line is nil except in so far as it may affect the sign of the voltage and current. In practice these conditions can, of course, be only approximated. So far as communication work is concerned, such conditions are neither desirable nor in evidence. We merely mention them here for the sake of general interest. Properties such as these are also exhibited by certain lumped-constant networks at properly chosen frequencies. These were first described by Boucherot and are sometimes named after him.

14. Effect of attenuation. In actual telephone lines and cables, the variation in voltage and current along the line is quite materially influenced by the presence of resistance and leakage, so that the discussion in the last four sections can serve only to indicate the general trend of these functions. In order to illustrate this, let us consider the following parameter values per mile which are fairly representative of an open telephone line consisting of a pair of 104 mil copper wires with 18-inch spacing.[1] Here

$$R = 10; \; L = 0.004; \; G = 10^{-6}; \; C = 0.008 \cdot 10^{-6}.$$

[1] The 18-inch spacing is for a so-called pole-pair, i.e., that pair of wires which straddles the pole. The parameters for a non-pole-pair do not differ appreciably from these.

The frequency at which these values hold corresponds to $\omega = 5{,}000$ radians per second. Using these in (43*a*) we find

$$\alpha = 0.00724 + j\,0.029 \tag{104}$$

i.e.,

$$\left.\begin{aligned} \alpha_1 &= 0.00724 \\ \alpha_2 &= 0.029 \end{aligned}\right\}. \tag{104a}$$

For the characteristic impedance we have by (53)

$$Z_0 = 748\ \underline{|12.5°}. \tag{105}$$

Using (78) and 104*a*), the wave length becomes

$$\lambda = \frac{2\pi}{0.029} = 216.5 \text{ miles}, \tag{106}$$

while for the phase velocity we get

$$v = \frac{\omega}{\alpha_2} = \frac{5000}{0.029} = 172{,}400 \text{ miles per second} \tag{107}$$

which is slightly less than the velocity of light.

It is interesting to compare these values with those which would result if the resistance and leakage were zero. Strictly speaking the inductance would become somewhat less if the resistance were zero because then the skin effect would be complete; but we shall neglect this small change. Then we get

$$\left.\begin{aligned} \alpha &= j\,0.0283 \\ Z_0 &= 707 \\ \lambda &= 222 \\ v &= 176{,}600 \end{aligned}\right\}. \tag{108}$$

Except for the fact that the attenuation is zero, these values do not differ very much from those for the actual line. The characteristic impedance shows perhaps the most significant change, but this will ordinarily not affect the line behavior to a great extent. This may be seen if we examine the effect which a change in Z_0 has upon the reflection coefficient r_R. Suppose we let

$$Z_0' = Z_0(1 + \Delta), \tag{109}$$

where Δ denotes the decimal change in Z_0. Then the new reflection coefficient becomes

$$r_R' = \frac{Z_R - Z_0(1 + \Delta)}{Z_R + Z_0(1 + \Delta)}. \tag{110}$$

Expanding and retaining only linear terms, we find for this the following approximate expression

$$r_R' = r_R - \frac{2\,Z_R Z_0 \Delta}{(Z_R + Z_0)^2} = r_R - \Delta', \tag{110a}$$

where r_R is the value for $\Delta = 0$. The effect upon the reflection coefficient is expressed by the second term in (110a). This is shown plotted versus Z_R in Fig. 9, from which we see that the change in the reflection coefficient has a maximum value at $Z_R = Z_0$, and that this maximum change equals half the decimal change in Z_0. If the latter were, say, 0.1, then the maximum variation that this could produce in r_R would be 0.05. Note, however, that this is the actual change in r_R, not the percentage change which would be infinite for $Z_R = Z_0$

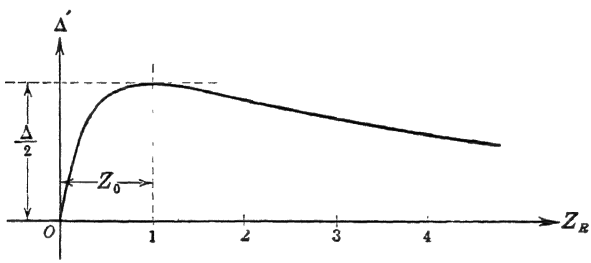


FIG. 9.—Change in the value of the reflection coefficient r_R as a function of the terminal impedance.

because then $r_R = 0$. For values of Z_R not in the vicinity of Z_0, the effect upon r_R is much less and hence for most purposes negligible.

Thus the outstanding effect in the dissipative case, at least so far as the behavior at a single frequency is concerned,[1] is the presence of attenuation. In our numerical example, (104a) gives the attenuation per mile. For a wave length the attenuation factor would be by (106)

$$216.5 \times 0.00724 = 1.568 \text{ napiers}$$

so that the ratio of the magnitudes of a single wave component at intervals of a wave length would be

$$e^{1.568} = 4.8,$$

i.e., the magnitude of either the incident or reflected wave decreases almost 80 per cent as it travels a distance of one wave length or 216.5 miles. Fig. 10 shows the variation in the net voltage over 111 miles (one-half wave length for the corresponding non-dissipative case) for a terminal resistance of 2,121 ohms (this makes $r_R = 0.5$ for neglected

[1] When the behavior is studied as a function of frequency, other very significant effects will be found.

R and G). Since this plot is drawn for the ratio E/E^+, the actual voltage is found by multiplying by E^+ which must be determined from (64). In the same figure is also shown the ellipse which is obtained for the corresponding non-dissipative case. A comparison of this ellipse and the polar curve for the actual line will give the reader some idea of the effect which the attenuation has upon the line behavior. The effect upon the phase function is evidenced by the fact that the vectors for $x = 0$ and $x = l$ are not exactly horizontal. The deviation, however, is not very great.

The student should draw the corresponding polar plot for the current as well as those for a purely reactive terminal impedance.

15. General circuit parameters. Very often in communication work we are not so much interested in the variation of voltage and current

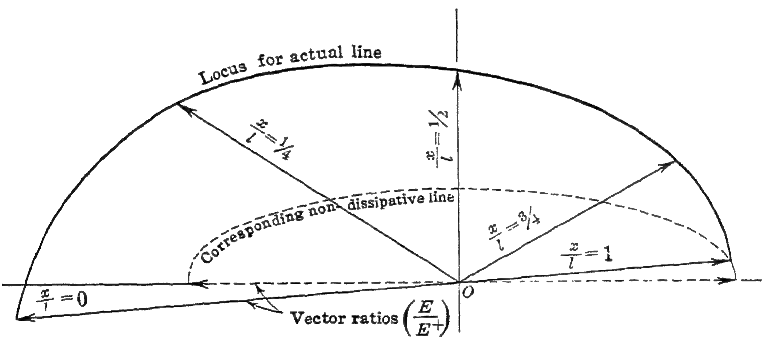


FIG. 10.—Polar plots showing the relative effect of dissipation.

along the line as we are in the relations between the sending- and receiving-end quantities alone. Hence it is convenient to have mathematical expressions which involve only such quantities. These may most conveniently be obtained from the hyperbolic forms (60d) and (61d). For example, if in the first of these we let $x = l$, we find

$$E_R = \frac{Z_0 Z_R E_g}{Z_0(Z_S + Z_R)\cosh \alpha l + (Z_S Z_R + Z_0^2)\sinh \alpha l}; \tag{111}$$

and if we use this relation together with

$$I_R Z_R = E_R, \tag{112}$$

the equations (60d) and (61d) may be expressed in terms of E_R and I_R, thus

$$\left.\begin{aligned} E(x) &= E_R \cosh \alpha(l - x) + Z_0 I_R \sinh \alpha(l - x) \\ I(x) &= \frac{E_R}{Z_0} \sinh \alpha(l - x) + I_R \cosh \alpha(l - x) \end{aligned}\right\}. \tag{113}$$

In particular, for $x = 0$ we have

$$\left.\begin{aligned} E_S &= E_R \cosh \alpha l + Z_0 I_R \sinh \alpha l \\ I_S &= \frac{E_R}{Z_0} \sinh \alpha l + I_R \cosh \alpha l \end{aligned}\right\}, \tag{114}$$

which express the sending- in terms of the receiving-end quantities directly. In order to do the reverse, we have merely to solve this set simultaneously for E_R and I_R in terms of E_S and I_S. This gives

$$\left.\begin{aligned} E_R &= E_S \cosh \alpha l - Z_0 I_S \sinh \alpha l \\ I_R &= -\frac{E_S}{Z_0} \sinh \alpha l + I_S \cosh \alpha l \end{aligned}\right\}. \tag{115}$$

When applying the results (114) and (115), the reader should recall the conventions as to positive directions for voltage and current, otherwise erroneous conclusions may be drawn. Fig. 11 illustrates these conventions schematically.

Positive flow of current is from left to right, and a positive voltage is one which rises from the datum plane to the conductor in the single-line diagram. The fact that the currents have this left-to-right direction causes the minus signs which appear in (115) but which are absent in (114). If the current directions are reversed, the signs in (115) will all become positive, but corresponding negative signs will appear in (114). Except for this matter of signs, the relations (114) and (115) form a symmetrical set; i.e., we may obtain one pair from the other by simply interchanging the subscripts S and R. This is to be expected from the fact that the line is a symmetrical system with respect to its two ends.

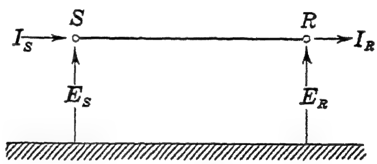


FIG. 11.—Conventions relative to positive directions for voltage and current.

The coefficients in the equations (114) are called the **general circuit parameters** and are usually denoted by

$$\left.\begin{aligned} \mathcal{A} &= \cosh \alpha l \\ \mathcal{B} &= Z_0 \sinh \alpha l \\ \mathcal{C} &= \frac{1}{Z_0} \sinh \alpha l \\ \mathcal{D} &= \cosh \alpha l \end{aligned}\right\}. \tag{116}$$

In terms of these, we have

$$\left.\begin{aligned} E_S &= \mathcal{A}E_R + \mathcal{B}I_R \\ I_S &= \mathcal{C}E_R + \mathcal{D}I_R \end{aligned}\right\}, \tag{114a}$$

and

$$\left.\begin{aligned} E_R &= \mathcal{D}E_S - \mathcal{B}I_S \\ I_R &= -\mathcal{C}E_S + \mathcal{A}I_S \end{aligned}\right\}. \tag{115a}$$

Mathematically, the pair of relations (114a) is spoken of as a linear transformation of the quantities E_R and I_R into E_S and I_S. The following arrangement of coefficients

$$\begin{Vmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{Vmatrix} \tag{117}$$

is called the *transformation matrix*, and

$$\begin{vmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{vmatrix} \tag{118}$$

the *determinant of the transformation*. The relations (115a) represent the inverse transformation, and

$$\begin{Vmatrix} \mathcal{D} & -\mathcal{B} \\ -\mathcal{C} & \mathcal{A} \end{Vmatrix} \tag{117a}$$

the *inverse* matrix to (117). In Chapter IV we shall show how the behavior of any linear four-terminal network may be analyzed in terms of its transformation matrix. The interconnection of several systems may be most effectively treated in this way.

Note that the transformation determinant (118) has the value

$$\mathcal{A}\mathcal{D} - \mathcal{B}\mathcal{C} = 1 \tag{119}$$

as may be easily verified by the use of (116). From the latter we also see that in this case

$$\mathcal{A} = \mathcal{D}. \tag{120}$$

Thus we see that, of the four general circuit parameters $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, only two are independent. We shall find later that the relation (119) holds generally for all linear four-terminal networks, and that, in addition, (120) holds for all those which are symmetrical with respect to their input and output terminals.

16. Input impedance. It is sometimes necessary to determine the impedance which the line presents at various points with an arbitrary load impedance. This may easily be done in terms of either the exponential forms (60a) and (61a), or the hyperbolic forms (60e) and

(61e) involving the equivalent line angles (83). Thus we get for the impedance looking into the line at any point x

$$\left.\begin{aligned}\frac{E(x)}{I(x)} = Z(x) &= Z_0 \coth\left[\alpha(l - x) - \rho_R\right] \\ &= Z_0 \frac{e^{\alpha(l-x)} + r_R e^{-\alpha(l-x)}}{e^{\alpha(l-x)} - r_R e^{-\alpha(l-x)}}\end{aligned}\right\}. \tag{121}$$

For the impedance looking in at the sending end, we have merely to put $x = 0$. Here two special cases are of interest. Namely, for $Z_R = 0$ we have $r_R = -1$, and by the second relation (121)

$$Z(0) = Z_{sc} = Z_0 \tanh \alpha l, \tag{122}$$

while for $Z_R = \infty$, $r_R = +1$, so that

$$Z(0) = Z_{oc} = Z_0 \coth \alpha l. \tag{123}$$

Z_{oc} and Z_{sc} are called the *open-circuit* and *short-circuit* input impedances of the line respectively. In terms of these the propagation function and characteristic impedance may be expressed. By forming the product or quotient of (122) and (123) we find, respectively

$$Z_0 = \sqrt{Z_{oc} \cdot Z_{sc}}, \tag{124}$$

and

$$\alpha l = \tanh^{-1}\sqrt{\frac{Z_{sc}}{Z_{oc}}} = \frac{1}{2} \ln \left(\frac{\sqrt{\dfrac{Z_{oc}}{Z_{sc}}} + 1}{\sqrt{\dfrac{Z_{oc}}{Z_{sc}}} - 1} \right). \tag{125}$$

These relations may be used to determine Z_0 and α experimentally from measurements of Z_{oc} and Z_{sc}. It is rather interesting that the characteristic impedance is determined by the product, and the propagation function by the quotient, of these impedances.

The student should note that the line itself is completely characterized by two functions no matter what form the characterization may take. So far we have seen that the line performance is expressible in terms of Z_0 and α, or any two of the general circuit parameters discussed in the previous section, or in terms of Z_{oc} and Z_{sc} as is evident from (124) and (125). The fact that, in any case, *two* functions suffice to describe the line characteristics uniquely, is fundamental.

17. Voltage, current, and power ratios. Very often we are not so much interested in the actual values of voltage, current, or power received at the far end of a line as we are in the ratio of output to input or its inverse. This ratio as a function of frequency is a direct indication of the quality of a given transmission facility.

Using the voltage equation (60a), we get for $x = l$

$$\frac{E_R}{E_g} = \frac{Z_0(1 + r_R)}{(Z_S + Z_0)(e^{\alpha l} - r_S r_R e^{-\alpha l})}. \tag{126}$$

For purposes of discussion, this may be put in a more effective form. Using (62) and (63) we find that

$$\frac{Z_0(1 + r_R)}{(Z_S + Z_0)} = \frac{Z_R}{Z_S + Z_R} \cdot (1 - r_S r_R), \tag{127}$$

so that we have

$$\frac{E_g}{E_R} = \frac{Z_S + Z_R}{Z_R} \cdot e^{\alpha l} \cdot (1 - r_S r_R)^{-1} \cdot (1 - r_S r_R e^{-2\alpha l}). \tag{128}$$

This is the desired ratio of input to output voltage. The reason for separating it into factors in the above manner is for convenience in discussion and calculation. The significance of the first factor, which involves only the sending- and receiving-end impedances, will be brought out shortly. The second or exponential factor is the most important because it usually contributes the largest share to the magnitude of the voltage ratio. Its value is determined entirely by the propagation function. The last two factors indicate the effect upon the ratio which is due to the discrepancies between the terminal and the characteristic impedances. If either $Z_S = Z_0$ or $Z_R = Z_0$, these factors both become unity. Their effect is usually small as compared to that of the exponential factor. Furthermore, the effect of the second-last factor is considerably greater than that of the last, particularly if the attenuation function α_1 is appreciable. This may be seen from the fact that the $r_S r_R$ -product in the last factor is multiplied by $e^{-2\alpha l}$, which is a very small number if the system is at all long or the attenuation relatively large. Thus the factors on the right-hand side of (128) are seen to be arranged in the order of their importance in affecting the net result. For a first approximation the first two will usually suffice. If this value is to be improved by partially taking the discrepancies between Z_S, Z_R, and Z_0 into account, the third factor may be included. The last factor then adds a final refinement to the numerical result. This is the situation in most practical cases.

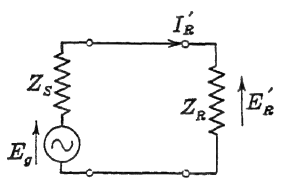


FIG. 12.—Source and load directly connected. Comparison with behavior for network of Fig. 4 places effect of line in evidence.

Although the ratio (128) is useful in many ways, it really does not specifically indicate the effect which the transmission line itself has

upon the system as a whole. In order to see this effect, we must form the ratio of that voltage which would be received if Z_S and Z_R were directly connected, to the voltage which actually appears. Fig. 12 illustrates the direct-connected situation. Here the effect of the leads between Z_S and Z_R is assumed to be negligible. Current and voltage at the receiving end are denoted by I_R' and E_R' respectively. Under these conditions we obviously have

$$\frac{E_g}{E_R'} = \frac{Z_S + Z_R}{Z_R}, \tag{129}$$

which illustrates the significance of the first factor in (128). From these two relations it now follows that

$$\frac{E_R'}{E_R} = e^{\alpha l} \cdot (1 - r_S r_R)^{-1} \cdot (1 - r_S r_R e^{-2\alpha l}). \tag{130}$$

This ratio shows the effect of inserting the transmission line between the terminal impedances Z_S and Z_R and is, therefore, called the *insertion ratio.*

In a similar way we find from (61*a*) the following ratio of sending- to receiving-end current:

$$\frac{I_g}{I_R} = \frac{I_S}{I_R} = e^{\alpha l} \cdot (1 - r_R)^{-1} \cdot (1 - r_R e^{-2\alpha l}), \tag{131}$$

in which the order of importance of the factors again tapers from left to right.

For the ratio of generator voltage to received current, we have merely to recall that

$$I_R Z_R = E_R, \tag{132}$$

and thus

$$\frac{E_g}{I_R} = \frac{Z_R E_g}{E_R}, \tag{133}$$

so that this is immediately obtained from (128).

Since

$$\left.\begin{aligned} I_R' Z_R &= E_R' \\ \text{and} \qquad I_R Z_R &= E_R \end{aligned}\right\}, \tag{134}$$

we see that

$$\frac{I_R'}{I_R} = \frac{E_R'}{E_R}. \tag{135}$$

Hence the insertion ratio for the current is the same as for the voltage.

In comparing the input-to-output voltage ratio (128) with (131) for the current, the reader will note a lack of similarity in these forms.

This is due to the fact that although the generator current equals the sending-end current, the same is not true for the voltage. In other words, the forms (128) and (131) are not on a comparable basis. The voltage ratio comparable to (131) is that for the sending- to receiving-end voltage. This may be found either directly from (60a) or more easily by setting $Z_S = 0$ in (128). The reader should convince himself of this by carrying out the former and comparing with the result of the latter, which is

$$\frac{E_S}{E_R} = e^{\alpha l} \cdot (1 + r_R)^{-1} \cdot (1 + r_R e^{-2\alpha l}). \tag{128a}$$

This is the voltage ratio which is on the same basis with (131). We see that neither depends upon Z_S.

In section 6 we saw that, whenever the line is terminated in its characteristic impedance, the voltage and current maintain the same phase relationship at all points in the system. In practice this condition is usually fairly nearly realized, so that the magnitude of the volt-ampere ratio then represents the corresponding power ratio.

For sending to receiving end, the volt-ampere ratio is given by the product of (131) and (128a), thus

$$\frac{E_S I_S}{E_R I_R} = e^{2\alpha l} \cdot (1 - r_R^2)^{-1} \cdot (1 - r_R^2 e^{-4\alpha l}). \tag{136}$$

For the volt-ampere insertion ratio we have by (130) and (135)

$$\frac{E_R' I_R'}{E_R I_R} = e^{2\alpha l} \cdot (1 - r_S r_R)^{-2} \cdot (1 - r_S r_R e^{-2\alpha l})^2. \tag{137}$$

It should be remembered that the magnitudes of these expressions become power ratios only when the line is terminated in Z_0, in which case we obviously get:

$$\frac{E_S}{E_R} = \frac{I_S}{I_R} = \sqrt{\frac{E_S I_S}{E_R I_R}} = \sqrt{\frac{E_R' I_R'}{E_R I_R}} = e^{\alpha l}. \tag{138}$$

In practice it has been found that it is more convenient to work with the logarithms of the above ratios than with the numerical ratios themselves. There are several reasons for this. We saw that in general the voltage, current, and volt-ampere ratios consist of products of various factors. The logarithm of such a product may be written as the sum of logarithms of the individual factors involved. This form evidently separates the contributions of the various factors in an additive manner and thus emphasizes their individual effects more strik-

ingly. For example, taking the natural logarithm of the voltage insertion ratio (130) we have

$$\ln\left(\frac{E_R'}{E_R}\right) = \alpha l - \ln(1 - r_S r_R) + \ln(1 - r_S r_R e^{-2\alpha l}). \qquad (130a)$$

Here the magnitudes of the successive terms usually taper from left to right in the manner already pointed out. Since the ratio itself is in general complex, the terms on the right of (130a) are also complex. The sum of the real parts indicates the net attenuation; the sum of the imaginary parts equals the net phase displacement. Separation of the first term in this way is very simple, and often gives the major contribution to both attenuation and phase shift, namely

$$\ln\left(\frac{E_R'}{E_R}\right) = \alpha_1 l + j\alpha_2 l \qquad \text{(approximately)}. \qquad (130b)$$

If we are interested only in the attenuation, as we frequently are, then we need consider only the logarithm of the absolute value of the ratio. Indicating absolute values by enclosing the corresponding complex quantities between vertical lines, we have

$$\ln\left|\frac{E_R'}{E_R}\right| = \alpha_1 l - \ln|1 - r_S r_R| + \ln|1 - r_S r_R e^{-2\alpha l}|, \qquad (130c)$$

which is spoken of as the *insertion loss.*

Similar expressions apply to the other ratios discussed above. The absolute values of these ratios are called **loss ratios,** and their logarithms are simply spoken of as **losses.** The separate terms, such as those on the right-hand side of (130c), are sometimes given separate names. In fact, the separation into terms is frequently carried still farther than has been done here, but there is nothing in this process which is of fundamental interest. The term $\alpha_1 l$ may be called the **attenuation loss** since it depends upon the attenuation function of the line alone. The rest are due to reflections at the line ends, and the process of designating them by names[1] is purely a matter of personal taste which may vary with the individual and hence is difficult to standardize.

If the line is dissipationless and properly terminated, the loss ratio is unity and its logarithm zero. Thus we say that the loss is zero. This again illustrates the advantage in working with the logarithmic ratios instead of the ratios themselves. Zero loss means a ratio whose absolute value is unity. A positive loss represents a loss ratio greater than unity, i.e., a decrease in the magnitude of the quantity involved.

[1] A fairly common usage is to call the second term in (130c), inclusive of its minus sign, the *reflection loss*, and the third term the *interaction loss* since it involves not only the reflection coefficients but also the propagation function.

If the loss ratio is less than unity, as it may be under certain conditions pointed out earlier, then the logarithm is negative. The loss is then said to be *negative,* or is also spoken of as a *gain.* The extremes of infinite and zero loss ratio become, in the logarithmic sense, infinite loss and infinite gain, respectively. The logarithmic function is thus seen to be more symmetrical with respect to loss and gain than the ratio itself.

There is still another reason why the logarithmic mode of expression is preferable. This has to do with the observed fact that the sensitivity of the human ear approximately obeys a square law. That is, a given sound will seem twice as loud when the pressure variation has reached the square of its original value. This is, of course, the same as to say that the logarithm of the original pressure variation has doubled. Thus the logarithmic ratio is a more natural way of expressing the intensity level of a received signal.

18. Attenuation units. In practical work it is convenient to have a unit in terms of which a given logarithmic ratio may be expressed, and to designate such a unit by an appropriate name. When the natural or Naperian system of logarithms is used (as has been done in all the above work) the logarithmic ratio is said to be expressed in **napiers.**[1] This name for the natural loss unit is chosen in honor of the inventor of the natural system of logarithms. Thus if the natural logarithm of a ratio equals, say, 2, we say that the loss for this ratio is two napiers. It happens, however, that the size of this natural unit is inconveniently large. In telephone work it has been found more convenient to express the loss in terms of the following definition involving the Briggs or common system of logarithms

$$\text{Loss} = 20 \log \left|\frac{E_1}{E_2}\right| = 20 \log \left|\frac{I_1}{I_2}\right|, \tag{139}$$

where "log" is written for the common system to distinguish it from "ln" which refers to the natural, and the subscripts 1 and 2 on E and I are used to indicate these ratios in more general terms. The factor 20 is arbitrarily inserted in order to give the unit a desired size. Historically the size of the unit was determined by the attenuation inherent in a mile of what in the early days was known as "standard cable." In that day the theory of transmission lines had not yet been developed to the point where engineers could make effective use of it, and the idea of logarithmic ratios was not available. When the theory became more thoroughly established, however, and a more useful definition for attenuation or loss was decided upon, the form (139) with the

[1] This is also written *neper,* which is said to be the original spelling of the name.

factor 20 was adopted because it led to a unit whose size most nearly approximated that for the old definition. The size of the latter was, of course, subject to a variation with frequency, but in the earlier days telephone lines were used only for voice frequencies, which were considered to be sufficiently well represented by an average value of 1,000 cycles per second. The loss in a mile of "standard cable" was understood to be referred to that frequency. The definition (139) is free from a frequency variation, and the unit size is still comparable to the old, so that engineers who are accustomed to thinking in terms of the old magnitude need not be inconvenienced by the change.

The original name for the unit defined by (139) was the **transmission unit,** which was abbreviated by the letters **T. U.** For properly terminated systems the relation (138) holds, and the magnitudes of the volt-ampere ratios become power ratios. We may then amplify the definition (139) and write

$$\left.\begin{aligned}\text{Number of T.U.} &= 20 \log \left|\frac{E_1}{E_2}\right| = 20 \log \left|\frac{I_1}{I_2}\right| \\ &= 10 \log \left|\frac{E_1 I_1}{E_2 I_2}\right| = 10 \log \frac{P_1}{P_2}\end{aligned}\right\}, \tag{139a}$$

where P_1 and P_2 are used to designate power.

More recently another unit was introduced and named in honor of Alexander Graham Bell. This unit was called the **bel** and defined by

$$\text{Number of bels} = \log \frac{P_1}{P_2}. \tag{140}$$

This definition, which leaves off the factor 10, obviously leads to a unit size which is ten times as large as the T.U. Since this size is inconvenient, it was decided to divide it into tenths and express loss in terms of tenths of bels or **decibels.** These, however, are again coincident in size with the transmission units; i.e.,

$$\text{Number of decibels} = \text{Number of T.U.}$$

The term decibel is abbreviated as **db**, and the final upshot of the matter is that db and T.U. amount to the same thing.

In order to determine the relation between napiers and decibels, we merely have to compare their respective definitions; for example

$$\text{Number of napiers} = \ln \left|\frac{E_1}{E_2}\right|,$$

while

$$\text{Number of db or T.U.} = 20 \log \left|\frac{E_1}{E_2}\right|.$$

If we recall that

$$\log x = \ln x \cdot \log e = 0.4343 \ln x,$$

then we have

$$\text{Number of db or T.U.} = 20 \times 0.4343 \ln \left|\frac{E_1}{E_2}\right|$$
$$= 8.686 \cdot \text{Number of napiers.} \tag{141}$$

Notice that the words "number of" are important. If we replace these by the words "size of," then the factor 8.686 appears on the other side of the equation, thus

$$8.686 \cdot \text{Size of a db or T.U.} = \text{Size of a napier.} \tag{141a}$$

In using such conversion equations the student should always be sure whether the words "number of" or "size of" are implied in the particular case in hand.

PROBLEMS TO CHAPTER II

2–1. An open telephone line has the parameter values: $R = 10.4$; $L = 0.00367$; $C = 0.00835 \cdot 10^{-6}$; $G = 0.8 \cdot 10^{-6}$. At the angular frequencies $\omega = 1000$; $\omega = 5000$; $\omega = 10{,}000$; $\omega = 20{,}000$, calculate the following:

(*a*) The attenuation function α_1.

(*b*) The phase function α_2.

(*c*) The wave length λ.

(*d*) The phase velocity v.

(*e*) The characteristic impedance Z_0 in magnitude and angle. Make sketches of these quantities versus frequency.

2–2. Repeat the calculations and plots indicated in Problem 2–1 for the cable having the following parameter values: $R = 83.2$; $L = 0.001$; $C = 0.062 \cdot 10^{-6}$; $G = 0.868 \cdot 10^{-6}$.

2–3. In Problems 2–1 and 2–2, neglect dissipation, and calculate and plot the indicated quantities. Compare with the corresponding dissipative values.

2–4. The open line of Problem 2–1 is considered at a frequency of 796 cycles per second and at this frequency is one wave length long. Determine the polar plots for the voltage and current for the following terminal conditions:

(*a*) $Z_S = Z_R = 700$.

(*b*) $Z_S = 700$; $Z_R = 2000$.

(*c*) $Z_S = 0$; $Z_R = \infty$.

(*d*) $Z_S = 0$; $Z_R = j\,1000$.

2–5. Repeat Problem 2–4 with the cable data of Problem 2–2 for the terminal conditions:

(*a*) $Z_S = Z_R = 500$.

(*b*) $Z_S = 500$; $Z_R = 2000$.

(*c*) $Z_S = 0$; $Z_R = 0$.

(*d*) $Z_S = 0$; $Z_R = j\,500$.

2–6. For the open line of Problem 2–1 neglect R and G and determine the ellipse diagrams for voltage and current for the terminal conditions specified in Problem 2–4. Compare with the polar plots obtained for that problem.

2–7. The open line of Problem 2–1 is 200 miles long and is terminated at either end in resistances of 700 ohms. Calculate the logarithmic insertion ratio (130*a*) for the angular frequencies $\omega = 1000$ to $\omega = 20{,}000$ at intervals sufficiently close to plot curves of the real and imaginary parts of each of the three terms in this expression as well as of their respective sums.

2–8. Repeat Problem 2–7 for the cable whose data are specified in Problem 2–2 for a length of 50 miles and terminal resistances of 500 ohms each.

2–9. A pair of No. 10 copper conductors spaced 8 in. on centers is excited from a source having a frequency corresponding to a wave length of 30 meters. For the terminal conditions $Z_S = 0$, $Z_R = \infty$; $Z_S = 0$, $Z_R = 0$, determine the standing wave patterns for voltage and current for lengths corresponding to the following fractions of a wave length: ¾, ⅞, 1. Neglect dissipation. Repeat for the line one wave length long but terminated at the far end in a pure inductance of one microhenry. Determine approximately the effect of line dissipation on these results.

2–10. Consider the line of Problem 2–9 one wave length long with $Z_S = 0$; $Z_R =$ 2000, 1000, 500 ohms pure resistance. For each case determine the ellipse diagrams for voltage and current. Neglect dissipation of the line.

2–11. The line of Problem 1–1 is directly connected to a source of negligible internal impedance, and is terminated at the far end in a pure inductance. If the line is one wave length long, how large should the terminal inductance be so that the maxima of the voltage standing waves will not exceed $\sqrt{2}$ times the maximum voltage of the source? If the terminal inductance is made larger than this value, will the maxima become larger or smaller?

2–12. In a Lecher wire system which is used for frequency measurement the high-impedance-indicating device is connected to the far end. A piece of copper wire is bridged across the line and moved gradually from the far end toward the source. When a distance of one-quarter wave length from the far end is reached, the indicator suddenly registers. Discuss the behavior of the system under this critical condition so as to show why it functions in this manner.

2–13. Utilizing the relations (114*a*), show that the impedance looking into a line terminated in Z_R may be expressed as

$$Z(0) = \frac{\mathcal{A}\, Z_R + \mathcal{B}}{\mathcal{C}\, Z_R + \mathcal{D}}.$$

Making use of the results of Problem 2–2, calculate the impedance looking into this cable at the frequencies indicated and for all combinations of the following lengths and terminal impedance values:

$Z_R = 500$ ohms pure resistance.

$Z_R = 2000$ ohms pure resistance.

$l = 50$ miles.

$l = 100$ miles.

2–14. A 100-mile length of open line as specified in Problem 2–1 is joined directly to 50 miles of cable as specified in Problem 2–2, and the latter is terminated in a resistance of 500 ohms. At the frequencies indicated in 2–1, calculate:

(*a*) The reflection coefficient at the end of the cable.

(*b*) The reflection coefficient at the end of the line.

(*c*) The ratio of sending-end voltage to that at the point where the line joins the cable.

(*d*) The overall voltage ratio.

CHAPTER III

PROPAGATION AND CHARACTERISTIC IMPEDANCE FUNCTIONS OF THE LONG LINE

1. Ideal behavior. From the discussion in the preceding chapter it should be clear that the quality of the transmission line as a communication facility depends entirely upon the functions α and Z_0 defined by equations (43*a*) and (53) respectively. According to these equations we see that the behavior of α and Z_0 versus frequency depends upon all four line parameters R, L, G, and C. It is important to study this dependence upon the line parameters so that we may be able to determine the most favorable combination of values for them, as well as to establish criteria for the readjustment of their values in order that the line behavior in a given case may meet a prescribed degree of quality. Before going into the details of this study, it is well to review briefly again the most desirable forms for these functions so that we may have them before us as a comparative standard.

In section 6 of the preceding chapter we saw that, whenever the line is terminated in its characteristic impedance, reflections are absent. This is a most desirable condition because reflections not only cause echoes, which are a particular form of combined phase and amplitude distortion,[1] but also give rise to such distortion even though the propagation function meets all the conditions for perfect transmission. Furthermore, it was pointed out in the first chapter that radiation in the steady state was due to, and took place in the vicinity of, the line ends, but that this phenomenon was wholly absent in the infinitely long line. When a line is terminated in its characteristic impedance, it is virtually infinitely long, so that the performance proceeds as though the termination did not exist. The greater the mismatch at a terminal, the greater will be the chance for the induction of electromotive forces in neighboring circuits. This may become noticeable at higher frequencies such as are used in carrier telephony. Transpositions are not effective in counteracting this phenomenon. Lastly, proper termination is desirable from the standpoint of transmission efficiency since under these conditions all the transmitted energy is accepted by the terminal device.

[1] This point will be discussed in Chapter XI.

From the practical standpoint, the condition of proper termination is most easily met when the characteristic impedance is a real constant. This will, incidentally, also make the transmitted energy a maximum, as may be seen independently by considering a source with an internal impedance Z_i feeding a load impedance Z_x. Assuming the voltage of the source unity, and writing

$$\left.\begin{aligned} Z_i &= R_i + jX_i \\ Z_x &= R_x + jX_x \end{aligned}\right\},$$

the power expended in Z_x becomes

$$P = \frac{R_x}{(R_i + R_x)^2 + (X_i + X_x)^2}.$$

Differentiating partially with respect to X_x and setting this equal to zero gives

$$X_x = -X_i.$$

Substituting this into the expression for P, the subsequent condition for maximum power with respect to R_x gives

$$R_x = R_i.$$

Thus if the two impedances are pure resistances, the condition for maximum power transfer is simply that they shall be equal. When they have reactive components, however, then the latter should be equal and opposite. This is equivalent to a resonance condition. It shows that when Z_i and Z_x are complex and equal, the power transfer is not a maximum. In the transmission line, Z_0 is the one impedance and the termination the other. The condition for no reflection can be realized simultaneously with that for maximum power transfer only when Z_0 reduces to a constant. Hence this is the ideal for the characteristic impedance function.

From (71) and (72) we see that if the line is terminated in Z_0 at both ends we have

$$E(x) = \frac{E_g}{2} \cdot e^{-\alpha x}, \tag{71a}$$

$$I(x) = \frac{E_g}{2Z_0} \cdot e^{-\alpha x}. \tag{72a}$$

Assuming a constant Z_0, the behavior is then entirely governed by α. When the steady state consists of a linear superposition of a number of harmonic components in the Fourier sense, then the distortionless transmission of the resulting periodic function obviously requires that the relative attenuation and the phase velocities for the individual

components all be alike. The requirement for no amplitude distortion will evidently be met if the real part of α is a constant, while the phase velocity according to (80a) will be independent of frequency if the imaginary part of α is a linear function of frequency. This defines the ideal behavior for the attenuation and phase functions.

The reader should note well that the conditions for a distortionless transmission system involve *both* α and Z_0. So long as the latter is not constant, the system will possess some amplitude and phase distortion even though α_1 be constant and α_2 a linear function of frequency. On the other hand, if the variation in Z_0 with respect to Z_S and Z_R is not very great, i.e., if the terminal conditions are reasonably good, then in any practical system the behavior is almost wholly governed by α. If the error involved in this assumption is of doubtful magnitude, the reflection effects must, of course, be calculated according to the methods already discussed.

So far the only distortionless case that we have said anything about is that for neglected dissipation. There $Z_0 = \sqrt{L/C}$, $\alpha_1 = 0$, and $\alpha_2 = \omega\sqrt{LC}$. In general, Z_0, α_1, and α_2 are rather complicated functions of frequency and in some cases differ very materially from their ideal behavior. With suitable readjustments in the line parameters, however, they may in general be made to assume their ideal characteristics to within a reasonable degree of error over any prescribed finite region of frequencies. Since for any one channel of communication only a finite frequency range is required, it is necessary only that the functions Z_0, α_1, and α_2 conform sufficiently well to their ideal standards within this range. If the neighboring regions are not utilized, it is needless to worry about the behavior of these functions there. In practice it is usually much simpler to readjust matters so that the behavior will be good over a narrow frequency range than it is to do this for a wide one. The location of the range in the frequency spectrum also has much to do with the practical difficulties involved. Whatever the case may be, the problem in general is to meet certain prescribed tolerances regarding amplitude and phase distortion over a given range of frequencies, and to do this with the minimum readjustments in the line parameters. To do more than is required would be a waste. The problem is to determine that set of readjustments which will just meet the requirements. This, at least, is the logical statement of the problem so far as the line itself is concerned. In present-day practice it is found more economical to correct for a large share of the shortcomings of the line by means of special networks which are inserted at the terminals. An appreciation of the economies involved in such practice requires first that the prob-

lem without auxiliary networks be understood. This we shall now undertake to present.

2. Separation of the propagation function. We shall begin with the discussion of the propagation function. From (43a) we have

$$\left.\begin{aligned} \alpha &= \sqrt{(R + jL\omega)(G + jC\omega)} \\ &= \alpha_1 + j\alpha_2 \end{aligned}\right\}. \tag{142}$$

Obviously the first problem is to separate the radical into its real and imaginary components. Squaring both sides of (142) we get

$$\alpha_1^2 - \alpha_2^2 + 2j\alpha_1\alpha_2 = (RG - LC\omega^2) + j\omega(LG + RC). \tag{143}$$

Equating reals and imaginaries, this gives

$$\left.\begin{aligned} \alpha_1^2 - \alpha_2^2 &= RG - LC\omega^2 \\ 2\alpha_1\alpha_2 &= (LG + RC)\omega \end{aligned}\right\}. \tag{144}$$

If we square each of these, add the results, and after some factoring again extract the square root, we find

$$\alpha_1^2 + \alpha_2^2 = \sqrt{(R^2 + L^2\omega^2)(G^2 + C^2\omega^2)}. \tag{145}$$

With this and the first equation (144) we can easily obtain α_1^2 and α_2^2 by addition and subtraction respectively. Thus we get

$$\alpha_1^2 = \tfrac{1}{2}\left[(RG - LC\omega^2) + \sqrt{(R^2 + L^2\omega^2)(G^2 + C^2\omega^2)}\right], \tag{146}$$

and

$$\alpha_2^2 = \tfrac{1}{2}\left[(LC\omega^2 - RG) + \sqrt{(R^2 + L^2\omega^2)(G^2 + C^2\omega^2)}\right]. \tag{147}$$

The square roots of these expressions formally at least represent the desired separation. It may be easily appreciated that the resulting forms are rather awkward from the standpoint of their discussion. For a given set of numerical values of R, L, G, and C, the process of plotting α_1 and α_2 versus frequency is, of course, a straightforward matter although somewhat long and tedious. What we wish to accomplish, however, is to discuss these functions analytically so that the relative effects of the various parameters will become more evident. In order to carry this out, various changes of variable as well as the introduction of several new parameters will be made. We hope that this process will not discourage the reader from continuing.

3. Discussion of the attenuation and phase functions. The primary object in the following manipulations is to put the analytic expressions for α_1 and α_2 in such a form that the departure from their ideal behavior in the general case becomes most evident. In attempting to carry out this idea we shall express these functions as products of their ideal values, and such factors which in the ideal case become unity.

Several substitutions are convenient in this connection. These are

$$\left.\begin{aligned} \frac{R}{L} &= a \\ \frac{G}{C} &= b \\ x &= \frac{\omega}{\sqrt{ab}} \end{aligned}\right\}. \tag{148}$$

Then (146) and (147) may be written

$$\alpha_1^2 = \frac{RG}{2}\left\{(1 - x^2) + \sqrt{\left(\frac{a}{b} + x^2\right)\left(\frac{b}{a} + x^2\right)}\right\} \tag{146a}$$

$$\alpha_2^2 = \frac{LC\omega^2}{2}\left\{(1 - x^{-2}) + \sqrt{\left(\frac{a}{b} + x^{-2}\right)\left(\frac{b}{a} + x^{-2}\right)}\right\}. \tag{147a}$$

The factors RG and $LC\omega^2$ are the ideal values referred to. The remaining factors indicate the departure from the ideal. The discussion which follows concerns itself chiefly with these remaining factors. Here the additional substitutions are useful

$$\left.\begin{aligned} m &= \tfrac{1}{2}\left(\sqrt{\frac{a}{b}} + \sqrt{\frac{b}{a}}\right) \\ n &= \tfrac{1}{2}\left(\sqrt{\frac{a}{b}} - \sqrt{\frac{b}{a}}\right) \end{aligned}\right\}. \tag{149}$$

With these we find that (146a) and (147a) become

$$\alpha_1 = \sqrt{RG} \cdot f_1(x) \tag{146b}$$

$$\alpha_2 = \omega\sqrt{LC} \cdot f_2(x) \tag{147b}$$

with

$$f_1(x) = \frac{1}{\sqrt{2}}\left\{1 - x^2 + \sqrt{1 + 2(m^2 + n^2)x^2 + x^4}\right\}^{1/2} \tag{150}$$

$$f_2(x) = \frac{1}{\sqrt{2}}\left\{1 - x^{-2} + \sqrt{1 + 2(m^2 + n^2)x^{-2} + x^{-4}}\right\}^{1/2}. \tag{151}$$

The functions $f_1(x)$ and $f_2(x)$ contain all that is pertinent to the variation of α_1 and α_2. They have several rather interesting properties. One of these is evident from (150) and (151), namely that

$$f_2(x) = f_1\left(\frac{1}{x}\right), \tag{152}$$

which states that the branch of $f_1(x)$ for the region $0 \leq x \leq 1$ is reciprocal to that of $f_2(x)$ for $1 \leq x \leq \infty$, and vice versa. Hence, if $f_1(x)$ is known for $0 \leq x \leq \infty$, then $f_2(x)$ may be found by reciprocation; or if $f_1(x)$ and

$f_2(x)$ are known in either of the regions $0 \leq x \leq 1$ or $1 \leq x \leq \infty$, then the balance of these functions may be got by geometrical construction.

The relationship between $f_1(x)$ and $f_2(x)$ goes even farther than this, as may be more readily seen if we make another change of variable, namely let

$$\left.\begin{aligned} y &= \frac{1}{2}\left(x - \frac{1}{x}\right) \\ x &= \sqrt{y^2 + 1} + y \end{aligned}\right\}. \tag{153}$$

Then we have

$$f_1(x) = x^{1/2}(\sqrt{m^2 + y^2} - y)^{1/2}, \tag{150a}$$

$$f_2(x) = x^{-1/2}(\sqrt{m^2 + y^2} + y)^{1/2}. \tag{151a}$$

From these we see that

$$f_1(x) \cdot f_2(x) = m, \tag{154}$$

which shows that $f_2(x)$ is the reciprocal of $f_1(x)$ with respect to the parameter m. Hence the branches of $f_1(x)$ and $f_2(x)$ are not only reciprocal about $x = 1$ for the regions $0 \leq x \leq 1$ and $1 \leq x \leq \infty$, but are also reciprocal about $f_1(1) = f_2(1) = \sqrt{m}$ within each region. This means that, if, for example, $f_1(x)$ is determined for $0 \leq x \leq 1$, then the balance of $f_1(x)$ and the whole of $f_2(x)$ can be found by geometrical construction alone!

Another aid toward plotting $f_1(x)$ and $f_2(x)$ is found if we differentiate (154) with respect to x. This gives

$$f_1(x) \cdot f_2'(x) + f_2(x) \cdot f_1'(x) = 0, \tag{155}$$

where primes denote differentiation with respect to the argument. If we note that

$$f_1(1) = f_2(1) = \sqrt{m} \tag{156}$$

we have for the point $x = 1$

$$f_2'(1) = -f_1'(1), \tag{155a}$$

i.e., the functions have equal and opposite slopes at this point. For $f_1(x)$ this slope is found most easily by considering twice the square of (150). Differentiating this with respect to x, we find

$$4f_1 \cdot \frac{df_1}{dx} = -2x + \tfrac{1}{2}(1 + 2(m^2 + n^2)x^2 + x^4)^{-1/2} \cdot (4x(m^2 + n^2) + 4x^3),$$

and setting $x = 1$, we have with the help of (156)

$$f_1'(1) = \tfrac{1}{2}(m^{1/2} - m^{-1/2}). \tag{157}$$

A further aid in plotting is found in series expansions of $f_1(x)$ and $f_2(x)$ about the points $x = 0$ and $x = \infty$. These will show how the

functions behave in the vicinity of the origin and infinity. Just a few terms will suffice to accomplish this. We find from (150) and (152) or (154)

$$\left.\begin{aligned} \text{About } x = 0\colon f_1(x) &= 1 + \frac{n^2x^2}{2} + \cdots \\ f_2(x) &= m\left(1 - \frac{n^2x^2}{2} + \cdots\right) \end{aligned}\right\}, \tag{158}$$

and

$$\left.\begin{aligned} \text{About } x = \infty\colon f_1(x) &= m\left(1 - \frac{n^2}{2x^2} + \cdots\right) \\ f_2(x) &= 1 + \frac{n^2}{2x^2} + \cdots \end{aligned}\right\}. \tag{159}$$

These show that the functions vary between the values 1 and m, and that they have zero slopes at the points $x = 0$ and $x = \infty$.

With the above information regarding $f_1(x)$ and $f_2(x)$, it is now a simple matter to plot these functions. This is done in Fig. 13, for which a

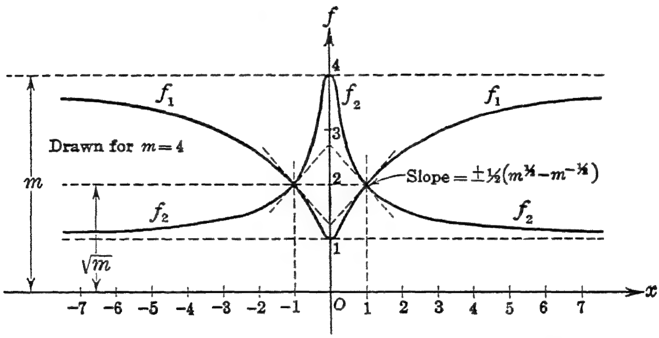


FIG. 13.—Functions showing departure from ideal behavior on the part of the attenuation and phase functions according to (146b) and(147b).

value $m = 4$ is arbitrarily chosen. The functions are plotted for negative as well as positive values of x in order to emphasize the fact that they are even functions. By (146*b*) and (147*b*) we, therefore, see that α_1 is an *even*, and α_2 an *odd*, function of frequency. Although negative frequencies have no physical significance, their consideration will be convenient later on in the transient analysis by means of the Fourier integral methods.

Finally in Fig. 14 we show plots of α_1 and α_2 corresponding to the plots of $f_1(x)$ and $f_2(x)$ in Fig. 13.

4. Amplitude and phase distortion. Assuming for the moment that the line is properly terminated, the functions α_1 and α_2 completely

control the amplitude and phase distortion for the system, while the functions $f_1(x)$ and $f_2(x)$ most conveniently indicate the magnitude of the distortion. If these functions were constants, the transmission would be distortionless. The amount of distortion in any case is best

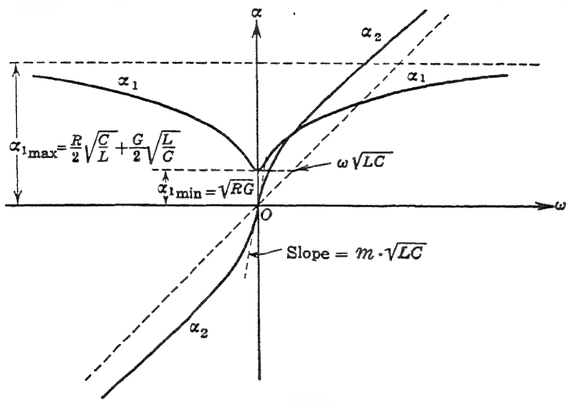


FIG. 14.—Plots of the attenuation and phase functions corresponding to the departures shown in Fig. 13.

expressed in terms of the percentage deviation of these functions from constant values over the essential frequency range for which the facility is to be used.

Fig. 13 shows that $f_1(x)$ has for its asymptote the value m, while $f_2(x)$ approaches unity with increasing frequency. It is also apparent that most of the distortion occurs at the lower frequencies; for frequencies above that corresponding to about $x = 4$ or 5 the distortion is certainly not appreciable. The maximum amount of distortion in any case evidently depends upon the value of the parameter m. If we examine its analytic

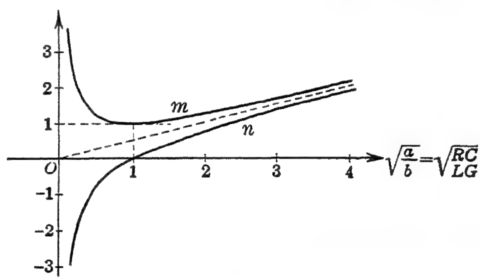


FIG. 15.—Dependence of the parameters (149) upon the line constants.

form as given by (149) we see that its value depends upon the ratio of a to b, which in turn are given by the line parameters, as expressed by (148). More precisely, m is equal to half the sum of $\sqrt{a/b}$ plus its reciprocal. Fig. 15 shows both the parameters m and n as functions of

this square root. We see that m has a minimum value of unity which occurs for

$$\sqrt{\frac{a}{b}} = 1$$

or

$$a = b,$$

which by (148) is equivalent to

$$\frac{R}{L} = \frac{G}{C}. \tag{160}$$

For all other values of $\sqrt{a/b}$ it is larger than unity. Referring to Fig. 13 again we see that if $m = 1$ the upper asymptote coincides with the lower and both f_1 and f_2 reduce to unity at all frequencies! Thus, when (160) holds, the attenuation and phase functions follow their ideal behavior. We shall see in a later section of this chapter that the condition (160) also reduces Z_0 to a real constant. This then is the condition for which a dissipative line will be distortionless. We saw earlier that the non-dissipative line was distortionless. Now we see that the dissipative line may also be distortionless, but only if the line parameters satisfy the condition (160).

In practice this condition never exists. The usual situation is that

$$\frac{R}{L} > \frac{G}{C}. \tag{161}$$

Theoretically there are obviously four ways in which the condition (161) may be made to approach that given by (160). We might try to decrease R, increase L, increase G, or decrease C. Practically there is no way of decreasing the capacitance, and so this method is out of the question. G might be increased by using a poorer insulator, but although the functions $f_1(x)$ and $f_2(x)$ could be made unity by a sufficient increase in G, equation (146*b*) shows that the resulting attenuation would thereby be increased. Hence this process cannot be recommended. Increasing L is a good remedy but rather difficult to carry out in practice owing to the fact that the increase in inductance must be uniformly distributed. One way in which this is accomplished is by uniformly wrapping the conductors with a ribbon made of some material having a high magnetic permeability. This is done in some ocean cables, and is referred to as **uniform loading**. As this is a very expensive process, however, it is not resorted to unless there is no other way of meeting the situation. On land lines the problem of increasing L is approximately met by inserting the added inductance in small concentrated amounts at regular intervals. In this way a uniform dis-

tribution may obviously be approximated, the approximation becoming better the smaller the intervals. This process is referred to as **lump loading.** Unless the intervals are small enough, however, this method may introduce serious distortion at the higher frequencies. Since the expense involved increases materially as the intervals are shortened, it is necessary to determine carefully the maximum interval that may be tolerated for a given case. This determination, which involves some fundamental principles that we have not yet discussed, will be taken up in a later chapter.

Finally, by decreasing R we not only approach the distortionless condition, but also decrease the final value of the attenuation function α_1 as is evident from (146*b*). This is, therefore, the most desirable way of improving the situation. Since it involves a considerable increase in the quantity of copper needed for a given facility, it is quite a costly process. In cables, the increase in wire size reduces the number of pairs that can be placed in a given space. In congested areas this space factor definitely limits the extent to which this method of improvement may be carried. On open wire lines, on the other hand, the present tendency is to use more copper and omit line loading as a means for improvement. Loading practice is being confined more and more to cable facilities where the wire size is limited.

It should also be pointed out in this connection that in practice, line loading or increasing the line copper is not done primarily in order to attain or approximate the distortionless condition. The chief object is to reduce the attenuation at higher frequencies by reducing the asymptotic value of α_1. Noting that the asymptote of $f_1(x)$ is m, that for α_1, is, by (146*b*)

$$\sqrt{RG} \cdot m,$$

or using (148) and (149) this is

$$(\alpha_1)_{\max} = \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}}. \tag{162}$$

It is interesting to compare this maximum value of the attenuation function with its minimum value which occurs at zero frequency and obviously is

$$(\alpha_1)_{\min} = \sqrt{RG}. \tag{163}$$

A study of these maximum and minimum values as functions of the parameters R, L, and G brings out some useful facts regarding the relative merits of readjusting their magnitudes in attempting to approach the distortionless condition or reducing the average attenuation.

This we can very effectively illustrate by considering the following data which are approximately representative of an open telephone line. Here

$$R = 10; \; L = 0.004; \; G = 0.8 \cdot 10^{-6}; \; C = 0.008 \cdot 10^{-6}$$

per mile. Using these values in (162) and (163), the curves of Figs. 16, 17, and 18 are obtained for $(\alpha_1)_{\max}$ and $(\alpha_1)_{\min}$ as functions of R, L, and G respectively. In these figures, dotted ordinates are drawn at the points corresponding to the original parameter values, and again at those values for which the distortionless line condition is satisfied. It will be seen that the two curves become tangent at these points.

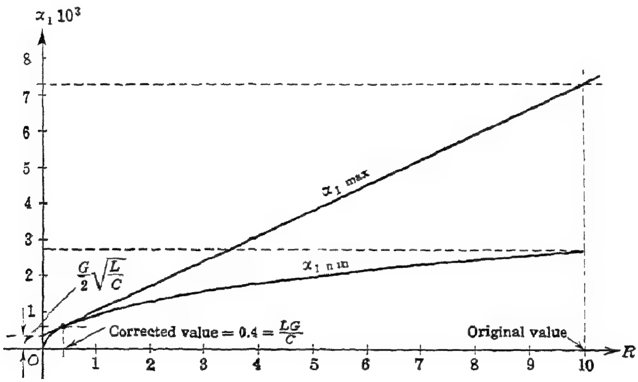


FIG. 16.—Dependence of the maximum and minimum values of attenuation upon the resistance parameter for constant values of L, G, and C.

The vertical separation at any point represents the maximum variation of the attenuation function with frequency for that parameter value.

Fig. 16 shows that decreasing R not only brings the two curves closer together, but also materially decreases the attenuation at both the low- and high-frequency limits. From Fig. 17 we see that increasing the inductance decreases the high-frequency value but leaves the minimum value unchanged. This method of approaching the distortionless condition is, therefore, not quite as desirable as already pointed out, although the considerable decrease in $(\alpha_1)_{\max}$ which is accomplished by loading is extremely useful in supporting the higher frequencies. In fact, line loading in the early days was chiefly heralded for this very reason. Fig. 17 brings out another very interesting point. We see that increasing L from its initial value is most effective throughout the first

few hundred per cent, say up to the value 0.03. Beyond this, very little is to be gained by a further increase, either in the reduction of $(\alpha_1)_{max}$ or in the attainment of the distortionless condition. Hence it would

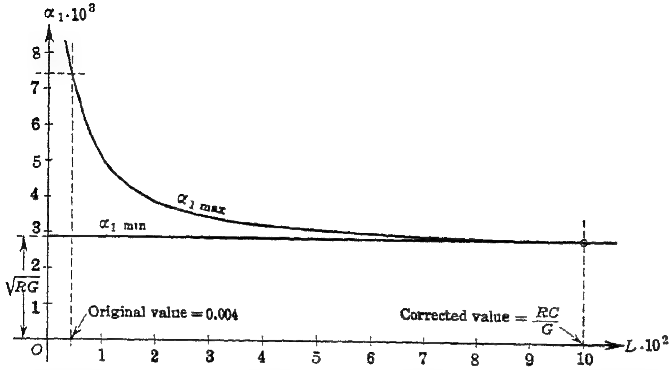


FIG. 17.—Dependence of the maximum and minimum values of attenuation upon the inductance parameter for constant values of R, G, and C.

hardly be economical to go beyond this value in any corrective procedure. Even a comparatively small amount of loading may thus do a very noticeable amount of good.

Fig. 18 shows how poor a policy it is to increase G in an attempt to

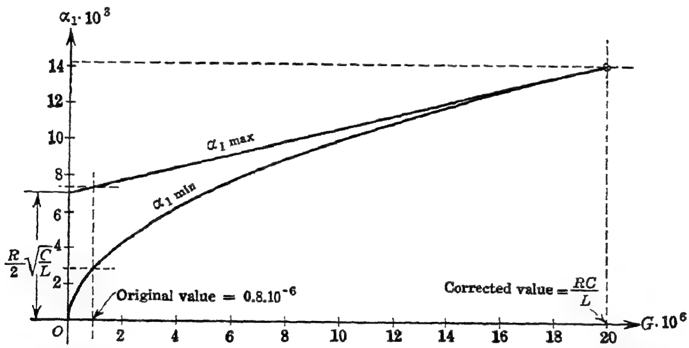


FIG. 18.—Dependence of the maximum and minimum values of attenuation upon the conductance parameter for constant values of R, L, and C.

approach the distortionless condition. Both the maximum and minimum values of the attenuation function are very materially increased by this procedure. The following table summarizes the situation with regard to the comparative results obtained from parameter readjustments.

	Parameter Values			Attenuation Values x 10^4		
	Original	30% Corrected	100% Corrected	Original	30% Corrected	100% Corrected
R	10	7.12	0 4	28.3–73.5	23.9–53.2	5.656
L	0.004	0.0328	0.1	"	28.3–32.8	28.3
G	0.8×10^{-6}	6.56×10^{-6}	20×10^{-6}	"	81.0–93.9	141.4

5. Tolerances and frequency limits. We have just seen that a considerable improvement with regard to amplitude and phase distortion may be secured with only a partial readjustment in the line parameters. In practice it is usually possible to tolerate a certain amount of distortion depending upon the quality of service required from the facility in question. In such a case, the maximum tolerance will be specified, and the question arises as to the extent to which the parameter values must be readjusted in order to meet this specification. A glance at the functions $f_1(x)$ and $f_2(x)$ of Fig. 13 shows that the frequency range for which the facility is to be used is also essential in deciding this question. Thus, for the same maximum tolerance, a range lying above a frequency corresponding to, say, $x = 4$, will evidently require less readjustment than one of the same width lying below this value. When both the frequency range and the maximum tolerance are specified, the practical problem admits of a unique solution.

In general the process of finding the solution is not so simple as it might be on account of the fact that it involves a cut-and-try procedure. It is simplified, however, by the fact that *the average variation in the attenuation function always equals the average variation* (from its ideal) *of the phase function.* This interesting point may be seen from the following. Let us denote the limits of the essential frequency range by x_1 and x_2. Then the average variation of the function $f_1(x)$ within this range is

$$\tfrac{1}{2}\{f_1(x_2) - f_1(x_1)\},$$

while its average value throughout the same range is

$$\tfrac{1}{2}\{f_1(x_2) + f_1(x_1)\},$$

so that the maximum percentage variation becomes

$$\frac{f_1(x_2) - f_1(x_1)}{f_1(x_2) + f_1(x_1)} \cdot 100\%. \tag{164}$$

Similarly the percentage variation of $f_2(x)$ within this same range is

$$\frac{f_2(x_2) - f_2(x_1)}{f_2(x_2) + f_2(x_1)} \cdot 100\%. \tag{165}$$

If we substitute the relation (154) into (165), we find that it reduces to (164) except for a reversal of sign. Thus the percentage amplitude and phase distortion defined on this basis are always the same in magnitude throughout any given frequency range!

When a maximum value for either of these variations is specified together with the essential frequency range, the following procedure may be adopted for determining the corresponding set of line parameters which will meet these specifications. Let us define the magnitude of the prescribed decimal variation in either $f_1(x)$ or $f_2(x)$ as

$$\delta_\alpha = \frac{f_1(x_2) - f_1(x_1)}{f_1(x_2) + f_1(x_1)}. \tag{164a}$$

From this we may easily get

$$\frac{1 - \delta_\alpha}{1 + \delta_\alpha} = \frac{f_1(x_1)}{f_1(x_2)}. \tag{166}$$

The relation between the variable x and the angular frequency is given by the last equation (148), namely

$$x = \frac{\omega}{\sqrt{ab}}. \tag{148a}$$

The limits of the prescribed frequency range may be denoted by ω_1 and ω_2 with the restriction that $\omega_1 \leq \omega_2$. In order to determine the corresponding values of x, the product ab must be known. Since this is one of the factors which we are trying to determine, we shall have to assume various values for it. This is more convenient if we let

$$\sqrt{ab} = k\sqrt{\omega_1\omega_2}, \tag{167}$$

and then assign values to the parameter k. The square root on the right-hand side of (167), incidentally, is the geometric mean frequency of the band. It is usually denoted as

$$\omega_0 = \sqrt{\omega_1\omega_2}. \tag{168}$$

Now we can write

$$x_1 = \frac{\omega_1}{\sqrt{ab}} = \frac{1}{k}\sqrt{\frac{\omega_1}{\omega_2}}, \tag{169}$$

and

$$x_2 = \frac{\omega_2}{\sqrt{ab}} = \frac{1}{k}\sqrt{\frac{\omega_2}{\omega_1}}. \tag{170}$$

For the sake of abbreviation let

$$s = \sqrt{\frac{\omega_1}{\omega_2}}. \tag{171}$$

Then

$$x_1 = \frac{s}{k};\ x_2 = \frac{1}{ks}. \tag{172}$$

Substituting these into (166) we have

$$\frac{1 - \delta_\alpha}{1 + \delta_\alpha} = \frac{f_1\left(\dfrac{s}{k}\right)}{f_1\left(\dfrac{1}{ks}\right)}. \tag{166a}$$

If we apply (152) and (154) to the right-hand side of this equation, we find

$$\frac{f_1\left(\dfrac{s}{k}\right)}{f_1\left(\dfrac{1}{ks}\right)} = \frac{f_1(ks)}{f_1\left(\dfrac{k}{s}\right)}, \tag{173}$$

from which we see that if a certain value of k satisfies (166a) then the reciprocal of this value will also satisfy it!

Since the form of the function $f_1(x)$ depends only upon the parameter m, as may best be seen from (150a), we may proceed to choose arbitrary values for k and graphically determine the corresponding m values which satisfy (166a). By the first relation (149), m determines the ratio of a to b, namely

$$\sqrt{\frac{a}{b}} = m \pm \sqrt{m^2 - 1}, \tag{174}$$

where the minus sign gives the reciprocal value. This together with (167) then determines a and b values which satisfy the prescribed conditions regarding tolerance and frequency limits.

Note that since k and its reciprocal both satisfy the same m and hence lead to the same a-to-b ratio, it is only necessary to choose k values either larger or smaller than unity. Furthermore, one choice of k and the correspondingly determined m will give us four pairs of a and b values which are the result of *either* of

$$\left.\begin{aligned} \sqrt{\frac{a}{b}} &= m_i + \sqrt{m_i^2 - 1} \\ \sqrt{\frac{a}{b}} &= m_i - \sqrt{m_i^2 - 1} \end{aligned}\right\} \tag{174a}$$

combined with *either* of

$$\left.\begin{aligned}\sqrt{ab} &= k_i\omega_0 \\ \sqrt{ab} &= \frac{\omega_0}{k_i}\end{aligned}\right\}, \tag{167a}$$

where k_i and m_i denote corresponding values.

The determination of m for a choice of k by means of (166a) is more conveniently accomplished if we make use of the function

$$F_1(x) = \frac{f_1(x)}{\sqrt{m}}. \tag{175}$$

Referring to Fig. 13 we see that $F_1(x)$ has the property that

$$F_1(1) = 1, \tag{176}$$

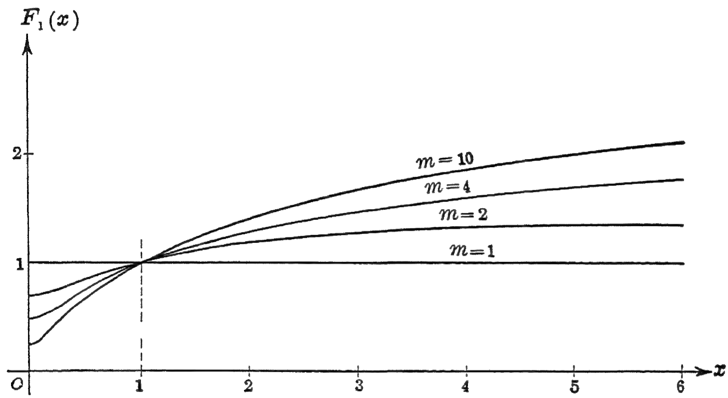


FIG. 19.—Normalized attenuation departure function, eq. (175), plotted against the frequency variable x, eq. (148), for various values of the parameter m, eq. (149).

i.e., a family of such curves for various m-values all pass through the point (1, 1). This makes a more easily usable chart. Fig. 19 shows such a family of curves.

Instead of (166a) we may, therefore, use

$$\frac{1-\delta_\alpha}{1+\delta_\alpha} = \frac{F_1(ks)}{F_1\left(\dfrac{k}{s}\right)}. \tag{166b}$$

For a given s, as determined by (171), the procedure is to choose successive k values, as for example

$$k_i = 1, 2, 3, \cdots.$$

With each of these the chart of Fig. 19 may be entered and correspond-

ing values of m_i found by trial which satisfy (166*b*) for the prescribed tolerance δ_α. The corresponding values of a and b are then found from (174*a*) and (167*a*).

With a sufficient number of these we may draw a set of curves of a versus b. Fig. 20 shows several such sets for various values of δ_α. We see that two curves, symmetrical about the 45° line, are obtained for each $\delta_\alpha \neq 0$, while the 45° line itself represents the condition for $\delta_\alpha = 0$. This might have been expected since the condition $a = b$ represents the distortionless line. The four pairs of values for a and b obtained from (174*a*) and (167*a*) for a single set of k_i and m_i give rise to four symmetrical points such as those marked 1, 2, 3, 4 in Fig. 20, two lying on one branch and two on the other.

The significance of the two branches for a given δ_α is that for each value of b there are *two* values of a which satisfy the prescribed data, one being larger and the other smaller than b. In any practical case we saw that for the original line or cable $a > b$, and the corrective procedure consists in decreasing a by either decreasing the resistance or increasing the inductance. This would, of course, be carried only so far as necessary. To carry the corrective procedure beyond the condition $a = b$ would be nonsense. Hence in practice only the curves of Fig. 20 lying *above* the 45° line are significant. Since b is usually fixed by the leakage and capacitance of the line, the solution to our corrective procedure becomes unique, although theoretically a *double infinity* of solutions present themselves.

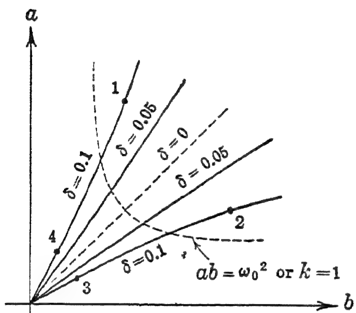


FIG. 20.—Ratio $a = R/L$ plotted against the ratio $b = G/C$ for various values of tolerance δ_α over a given frequency band.

For the sake of theoretical interest it might be added that the curves of Fig. 20 will in general be grouped more closely about the 45° line for wider frequency limits and for smaller values of the geometric mean frequency ω_0. For solutions corresponding to points below the hyperbola $k = 1$, the frequency band is located so that ω_0 lies *above* the point $x = 1$, while a point above this hyperbola places ω_0 *below* $x = 1$. Mathematically expressed this is the same as saying that for points below the hyperbola $k = 1$

$$\frac{\omega_0}{\sqrt{ab}} > 1,$$

while for those above this curve

$$\frac{\omega_0}{\sqrt{ab}} < 1.$$

For solutions lying on this hyperbola the frequency band is geometrically located about the frequency corresponding to $x = 1$.

In most practical cases conditions are such that approximations can be made which materially simplify the above general procedure. Usually the upper frequency limit ω_2 is high enough so that the function $f_1(x)$ may be assumed equal to its asymptotic value at this limit with sufficient accuracy. This means that in (166) we may let

$$f_1(x_2) = m. \tag{177}$$

For the value of the function at the lower limit it is sufficient to use the first two terms of the expansion (159). For values of δ_α less than 0.1, the error introduced by this procedure is not very great as may easily be verified. Consistent with these approximations for the right-hand side of (166), the left-hand side may likewise be put into an approximate form which is justified for values of δ_α less than 0.1.

Thus

$$\frac{1 - \delta_\alpha}{1 + \delta_\alpha} \cong (1 - \delta_\alpha)^2 \cong 1 - 2\delta_\alpha. \tag{178}$$

With these simplifications (166) becomes

$$1 - 2\delta_\alpha = 1 - \frac{n^2}{2x_1^2}. \tag{179}$$

Expressing this in terms of frequency by means of (148a) we have

$$\delta_\alpha = \frac{n^2 ab}{4\omega_1^2},$$

or by using the second relation (149) this is

$$\delta_\alpha = \frac{(a - b)^2}{16\omega_1^2}. \tag{180}$$

When the tolerance and the lower frequency limit are specified, the condition on the line parameters is expressed by

$$(a - b) = 4\omega_1\sqrt{\delta_\alpha}. \tag{180a}$$

Substituting from (148) we have

$$\frac{R}{L} - \frac{G}{C} = 4\omega_1\sqrt{\delta_\alpha}. \tag{180b}$$

If the ratio G/C is given, the necessary R/L ratio for which the variation δ_α will not be exceeded by either $f_1(x)$ or $f_2(x)$ for all frequencies above ω_1 may easily be found. As an illustration suppose we consider the open line represented by the parameter values on page 90 and stipulate that the inductance shall be increased until the maximum average variation δ_α does not exceed 0.05 for all frequencies in the voice range. Assuming that the lowest essential voice frequency is 100 cycles per second, we have

$$\omega_1 = 2\pi \cdot 100.$$

The upper frequency limit need not be considered since we have assumed it to be sufficiently high so that $f_1(x)$ very nearly equals the value m at this limit. The accuracy of this assumption may later be checked. We thus have

$$\left(\frac{R}{L}\right)_{\text{new}} - \frac{G}{C} = 8\pi \cdot 100 \cdot \sqrt{0.05} = 562,$$

and since for this line

$$\frac{G}{C} = 100,$$

we get

$$\left(\frac{R}{L}\right)_{\text{new}} = 662$$

for the required ratio of resistance to inductance. Since the given ratio is

$$\left(\frac{R}{L}\right)_{\text{old}} = 2500,$$

we see that the inductance must be increased until

$$L_{\text{new}} = 3.78 L_{\text{old}}.$$

In practice the introduction of additional inductance also increases the resistance. The process of taking this into account is simple and need not be taken up here. With this increase the inductance of the loaded line will be

$$L_{\text{new}} = 0.0151 \text{ henry per mile.}$$

If we assume that the upper limit of the voice frequency range is

$$\omega_2 = 2\pi \cdot 4000,$$

we find by substituting into (159) that

$$f_1(x_2) = m(1 - 0.000062)$$

which is certainly near enough to the value m. The result (180b) is, therefore, a convenient criterion. For larger values of δ_α it will indicate an R/L ratio somewhat smaller than necessary, but for most practical purposes its accuracy is sufficient. In cases of doubt, the error can, of course, be determined.

6. Phase and group velocity. In this section we wish to point out certain transmission properties of the long line which are brought out by a closer study of the phase function α_2. In the previous chapter we saw by equation (80a) p. 50 that the velocity of phase propagation, which is the velocity with which the waves travel in the harmonic steady state, is given by

$$v = \frac{\omega}{\alpha_2}. \tag{80a}$$

If we use (147b) this is

$$v = \frac{1}{\sqrt{LC} \cdot f_2(x)}. \tag{181}$$

In the foregoing we have seen that when the line is dissipationless, or satisfies the condition (160) which makes it distortionless, the function $f_2(x)$ becomes unity at all frequencies. Since in general $f_2(x)$ is greater than unity, we see that in the dissipationless or distortionless cases the phase velocity is constant and has its maximum value, which is

$$v_0 = \frac{1}{\sqrt{LC}}. \tag{182}$$

In terms of this value, and using the relation (154), we may write instead of (181)

$$v = v_0 \cdot \frac{f_1(x)}{m}. \tag{181a}$$

This clearly shows the variation of phase velocity with frequency in the general case. In fact the curve for $f_1(x)$ of Fig. 13 may be used to represent the velocity v by making a suitable change in the vertical scale in agreement with (181a). We thus see that the lower frequencies travel more slowly and that the value v_0 (which we recall is nearly equal to the velocity of light for open lines) is approached asymptotically as the frequency tends toward infinity. The ratio of this highest velocity v_0 to the lowest is, in the general case, equal to m.

This is the situation in the steady state when a single harmonic voltage is impressed upon the line. In communication work, however, we rarely find a situation as simple as this. In fact it may be questionable whether the average condition in practice may even be considered as

a steady state. However, we shall leave this question out of the argument until a later chapter. Meanwhile it is quite instructive as well as interesting to consider what happens when a steady state is composed of a linear superposition of a number of harmonic functions of different frequencies. Such an array of functions we shall call a **frequency group**. Furthermore, we shall confine our attention to a particular kind of frequency group, namely one for which all the amplitudes are equal and all the frequency differences equal.

Such a frequency group is illustrated in Fig. 21, which, incidentally, is called an **amplitude spectrum** representation for the group. The abscissae for this figure indicate the angular frequencies of the individual components; the ordinates indicate their respective amplitudes. This group is composed of one component whose frequency is the arithmetic mean, and n adjacent pairs equally spaced. The frequency interval is denoted as $\delta\omega$, and the "width" of the group is designated as

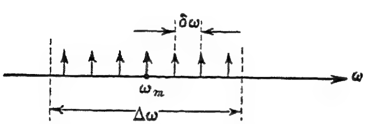


FIG. 21.—Amplitude spectrum of a uniform frequency group.

$$\Delta\omega = (2n + 1)\delta\omega. \tag{183}$$

It may seem to the reader that this group is rather special and, therefore, would hardly have any practical significance. However, an arbitrary distribution of amplitude versus frequency can always be approximately thought of as composed of a superposition of such groups as these. Such a representation for an arbitrary amplitude versus frequency distribution will become more exact the smaller the width $\Delta\omega$ of the individual groups is chosen. We will see later that by letting the group width approach zero, any time function may be accurately represented as a linear superposition of an infinite number of such groups. Thus the consideration of a group as is illustrated by Fig. 21 forms the basis for the exact analysis of the most general case.

To continue with the consideration of this frequency group, our first object is to show what the resulting wave picture for such an array of frequencies looks like. In order to do this we must write down the analytic expression for a group of waves whose frequencies and amplitudes correspond to the above distribution. If we assume for simplicity that the amplitudes are all unity, then the analytic expression for a component wave of angular frequency ω is

$$\cos{(\alpha_2 x - \omega t)}.^1$$

[1] The variable x here represents distance along the line as in Chapter II.

In particular for the *mean* frequency ω_m this is

$$\cos(\alpha_{2m}x - \omega_m t).$$

If we denote the corresponding change in α_2 to a frequency change $\delta\omega$ by $\delta\alpha_2$, then the next component wave in the group is given by

$$\cos[(\alpha_{2m} + \delta\alpha_2)x - (\omega_m + \delta\omega)t].$$

In order to simplify the writing of the entire group, let us introduce the notation

$$\left.\begin{aligned} \Omega &= \alpha_2 x - \omega t \\ \delta\Omega &= \delta\alpha_2 x - \delta\omega t \\ \Omega_m &= \alpha_{2m}x - \omega_m t \end{aligned}\right\}. \tag{184}$$

Then the wave group corresponding to the frequency group defined above becomes

$$f(x-vt) = \cos\Omega_m t + \left\{\begin{matrix} \cos(\Omega_m+\delta\Omega)+\cos(\Omega_m+2\delta\Omega)+ \cdots +\cos(\Omega_m+n\delta\Omega) \\ \cos(\Omega_m-\delta\Omega)+\cos(\Omega_m-2\delta\Omega)+ \cdots +\cos(\Omega_m-n\delta\Omega) \end{matrix}\right\}. \tag{185}$$

This finite sum of harmonic components may be evaluated by applying the trigonometric formulae for sums and differences of angles.[1] Thus (185) may be written

$$f(x - vt) = (2n + 1) \cdot F(\delta\Omega) \cdot \cos\Omega_m, \tag{185a}$$

where

$$F(\delta\Omega) = \frac{\sin\left[(2n+1)\dfrac{\delta\Omega}{2}\right]}{(2n+1)\cdot\sin\dfrac{\delta\Omega}{2}}. \tag{185b}$$

The result (185*a*) shows that the wave group may be visualized as consisting of the mean-frequency wave

$$\cos\Omega_m = \cos(\alpha_{2m}x - \omega_m t)$$

enclosed by an envelope whose form is determined by

$$(2n + 1) \cdot F(\delta\Omega).$$

This envelope is itself a wave; i.e., it has a fixed form but travels along the line as time increases. In order to get a better idea of what the shape of this envelope looks like, we will plot it versus distance for the instant $t = 0$.

[1] See for example, W. E. Johnson, "A Treatise on Trigonometry," Macmillan & Co., 1889, p. 320, e.s.

Then we have

$$F(\delta\Omega) = F(\delta\alpha_2 x) = \frac{\sin\left[(2n+1)\cdot\frac{\delta\alpha_2}{2}\cdot x\right]}{(2n+1)\cdot\sin\frac{\delta\alpha_2}{2}x}. \tag{186}$$

If the number of components in the group is large, i.e., if n is large, then for small values of x this is approximately given by

$$F(\delta\alpha_2 x) \cong \frac{\sin\left[(2n+1)\cdot\frac{\delta\alpha_2}{2}\cdot x\right]}{(2n+1)\cdot\frac{\delta\alpha_2}{2}\cdot x}, \tag{186a}$$

which is of the form

$$\frac{\sin u}{u}.$$

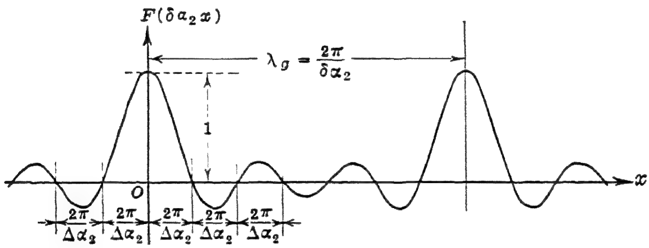


FIG. 22.—Wave envelope corresponding to the frequency group of Fig. 21.

This function crops up quite frequently in connection with network problems, and the student may as well familiarize himself somewhat with its general characteristics. The form (186*a*), of course, may be used to plot the function only in the vicinity of $x = 0$. The actual function (186) is obviously periodic in x; i.e., it repeats after a distance

$$\lambda_g = \frac{2\pi}{\delta\alpha_2}. \tag{187}$$

In Fig. 22 is shown a plot of (186) for $n = 3$. In this figure the quantity

$$\Delta\alpha_2 = (2n+1)\delta\alpha_2 \tag{188}$$

is used to denote the *width* of the *group* with respect to α_2 just as $\Delta\omega$ in (183) denotes the *width* with respect to ω.

Thus we see that the envelope which encloses the mean-frequency wave has the effect of grouping or "bunching" this wave into humps which come at intervals of λ_g miles. This distance, which is inversely

proportional to the increment $\delta\alpha_2$ and hence inversely proportional to the corresponding frequency increment $\delta\omega$, is called the **group wave length.** The length of any "ripple" in the envelope, which is half the length of the main hump, is

$$\frac{2\pi}{\Delta\alpha_2} = \frac{2\pi}{(2n+1)\cdot\delta\alpha_2}$$

and hence depends upon the "width" of the frequency group.

The complete picture of the wave group as it travels along the line is, therefore, as illustrated by Fig. 23, which is again drawn for $n = 3$. A further point of interest in connection with this wave group is now revealed by the fact that the velocity with which the envelope travels

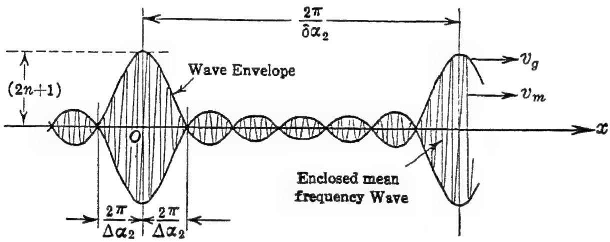


FIG. 23.—Resulting wave group corresponding to the frequency group of Fig. 21.

is in general not equal to that with which the enclosed mean-frequency wave travels. The latter is evidently

$$(v_{ph})_m = \frac{\omega_m}{\alpha_{2m}}. \tag{189}$$

Since the arguments of the sine-functions in (185*b*) involve the quantity

$$\delta\Omega = \delta\alpha_2 x - \delta\omega t,$$

a point of constant phase for the envelope will be given by

$$d(\delta\alpha_2 x - \delta\omega t) = \delta\alpha_2 \cdot dx - \delta\omega \cdot dt = 0,$$

from which

$$\frac{dx}{dt} = v_g = \frac{\delta\omega}{\delta\alpha_2}. \tag{190}$$

This is the velocity with which the *envelope* travels. If the width $\Delta\omega$ of the frequency group is appreciable, the value

$$\frac{\delta\omega}{\delta\alpha_2}$$

may vary somewhat throughout the group. In general, the group

width $\Delta\omega$ will be taken small enough so that this variation may be neglected, and we can write

$$v_g = \left(\frac{d\omega}{d\alpha_2}\right)_{\omega = \omega_m}. \tag{190a}$$

That is, the velocity of the envelope, which is also called the **group velocity**, is equal to the inverse slope of the α_2-versus-ω curve at the point $\omega = \omega_m$.

In the dissipationless or the distortionless cases for which

$$\alpha_2 = \omega\sqrt{LC},$$

we have

$$v_{ph} = v_g = \frac{1}{\sqrt{LC}},$$

i.e., both the *phase* and *group velocities* are constant and equal to each other. In general, however, the group and phase velocities will vary with frequency and will not equal each other.

In order to study the variation of the group velocity with frequency for the long line, we must first determine the variation of the derivative of α_2 with respect to ω. The expression for this derivative becomes quite involved if written entirely in terms of ω. However, if by the use of (148) and (147*b*) we first write

$$\alpha_2 = \sqrt{RG} \cdot x \cdot f_2(x),$$[1]

and then note that

$$\frac{d\alpha_2}{d\omega} = \frac{d\alpha_2}{dx} \cdot \sqrt{\frac{LC}{RG}},$$

so that

$$\frac{d\alpha_2}{d\omega} = \sqrt{LC} \cdot \frac{d}{dx}(x \cdot f_2(x)),$$

we get by using (151*a*) and recalling the substitution (153)

$$\frac{d\alpha_2}{d\omega} = \sqrt{LC}\left[\frac{(\sqrt{m^2 + y^2} + y)^{1/2} \cdot (\sqrt{m^2 + y^2} + y + \frac{1}{x})}{2\sqrt{x}\sqrt{m^2 + y^2}}\right]. \tag{191}$$

This form is quite convenient for plotting. Fig. 24 shows this derivative as a function of x for $m = 4$. The units for the vertical scale are equal to $\sqrt{LC}$, which is the dissipationless or distortionless value. We see that the function has a minimum value less than this, which is quite an interesting point. In order to locate the frequency at which this minimum occurs, we must equate the second derivative of α_2 to zero and solve for the corresponding x. This process becomes exceedingly

[1] Note that here x is the variable as defined by (148) and not distance.

involved, as may easily be appreciated. The simplest way of accomplishing it is not to attempt differentiating (191) again, but to go back to the original expression (142) for α. Since the derivative of the imaginary part equals the imaginary part of the derivative, we may accomplish the desired end by differentiating (142) twice with respect to ω,

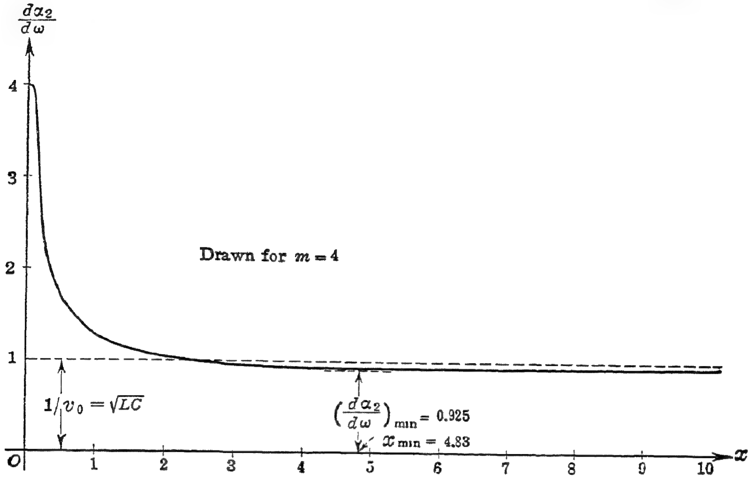


FIG. 24.—Delay time (units equal to $\sqrt{LC}$ seconds) of the uniform transmission line as a function of the frequency variable x, eq. 148, for a constant value of the parameter m, eq. 149, assuming constant line parameters.

and then equate the imaginary part of the result to zero. We shall omit the actual work here, and merely state the result, namely that the minimum of (191) occurs at

$$x_{\min} = \frac{\sqrt{m^2+3}+m}{\sqrt{3}}, \tag{192}$$

and that the minimum value found by substituting this back into (191) becomes

$$\left(\frac{d\alpha_2}{d\omega}\right)_{\min} = \sqrt{LC}\cdot\frac{\sqrt{3}}{4}\left\{\sqrt{\frac{m}{\sqrt{m^2+3}+m}}+\sqrt{\frac{\sqrt{m^2+3}+m}{m}}\right\}. \tag{193}$$

It is now interesting to compare the phase and group velocities as functions of frequency. They are respectively

$$\left.\begin{aligned} v_{ph} &= \frac{\omega}{\alpha_2} \\ v_g &= \frac{d\omega}{d\alpha_2} \end{aligned}\right\}. \tag{194}$$

Using (181a) for the former and inverting Fig. 24 for the latter, we find the curves of Fig. 25 for these two quantities, again drawn for $m = 4$. We see from this figure that in general the group velocity is *always* larger than the phase velocity! This means that when a wave group travels along a line, the velocity of the envelope, i.e., of the groups or humps, will in general be greater than that of the enclosed wave, and that these velocities merge toward the distortionless value v_0 as the mean frequency of the group tends toward infinity.

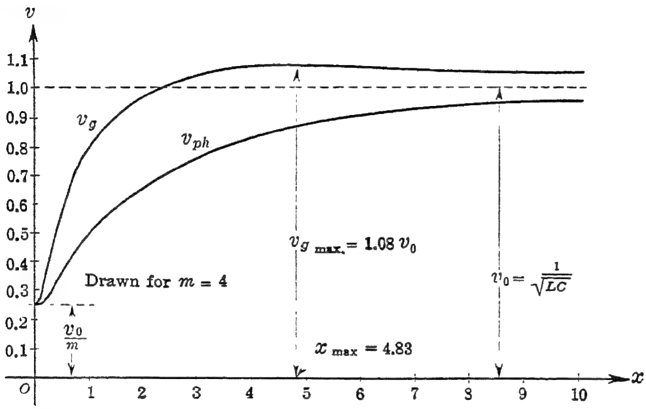


FIG. 25.—Phase and group velocities of the uniform transmission line as functions of the frequency variable x. These curves correspond to the conditions underlying Fig. 24.

From (193) we note that this expression has its smallest value for $m \to \infty$, and that this value is

$$\left[\left(\frac{d\alpha_2}{d\omega}\right)_{\min}\right]_{m\to\infty} = \sqrt{LC}\cdot\frac{3\sqrt{3}}{4\sqrt{2}}. \tag{195}$$

This means that the largest maximum value of v_g will occur for lines for which either

$$\frac{R}{L} >> \frac{G}{C}$$

or

$$\frac{G}{C} >> \frac{R}{L},$$

and that this largest value is

$$v_{g\,\max\,\max} = \frac{4\sqrt{2}}{3\sqrt{3}}\,v_0 = 1.088\,v_0. \tag{196}$$

7. Discussion of the characteristic impedance function. This function, like the propagation function, is also in general complex, and hence may be split into real and imaginary components. However, it is usually more convenient to study the characteristic impedance function in its polar form, i.e., in the form of magnitude and angle. Thus we write

$$Z_0 = \sqrt{\frac{R + jL\omega}{G + jC\omega}} = |Z_0| \underline{|\zeta} \tag{197}$$

and proceed to determine the functions $|Z_0|$ and ζ. For the magnitude we evidently have

$$|Z_0| = \left(\frac{R^2 + L^2\omega^2}{G^2 + C^2\omega^2}\right)^{1/4} = \sqrt{\frac{L}{C}} \cdot \left(\frac{a^2 + \omega^2}{b^2 + \omega^2}\right)^{1/4}, \tag{198}$$

where the relations (148) have been used. The quantity $\sqrt{L/C}$ is obviously the dissipationless value. The following parenthesis factor in (198), therefore, shows the effect of dissipation and hence contains all that is of interest in the study of $|Z_0|$. Let us write

$$|Z_0| = \sqrt{\frac{L}{C}} \cdot z_0, \tag{198a}$$

and then with the last relation (148) we have

$$z_0(x) = \left(\frac{\dfrac{a}{b} + x^2}{\dfrac{b}{a} + x^2}\right)^{1/4} \tag{199}$$

From this result we can obtain the interesting relation that

$$z_0(x) \cdot z_0\left(\frac{1}{x}\right) = \sqrt{\frac{a}{b}} \tag{200}$$

from which we note that when $z_0(x)$ has been plotted for the interval $0<x<1$, it may be extended into the interval $1<x<\infty$ by geometrical construction according to (200). From this relation it also follows that

$$z_0(1) = \sqrt[4]{\frac{a}{b}}, \tag{201}$$

as is also evident from (199).

In the vicinity of $x = 0$ the relation (199) may be expanded to obtain

$$\begin{aligned} z_0(x) &= \left(\frac{a}{b} + x^2\right)^{1/4}\left(\frac{b}{a} + x^2\right)^{-1/4} = \sqrt{\frac{a}{b}}\left(1 + \frac{bx^2}{a}\right)^{1/4}\left(1 + \frac{ax^2}{b}\right)^{-1/4} \\ &= \sqrt{\frac{a}{b}}\left(1 + \frac{bx^2}{4a} + \cdot\cdot\cdot\right)\left(1 - \frac{ax^2}{4b} + \cdot\cdot\cdot\right) \\ &= \sqrt{\frac{a}{b}}\left\{1 - \frac{x^2}{4}\left(\frac{a}{b} - \frac{b}{a}\right) + \cdot\cdot\cdot\right\}, \end{aligned} \tag{202}$$

while for large values of x we may write

$$\begin{aligned} z_0(x) &= \left(1 + \frac{a}{bx^2}\right)^{1/4}\left(1 + \frac{b}{ax^2}\right)^{-1/4} \\ &= \left(1 + \frac{a}{4bx^2} + \cdot\cdot\cdot\right)\left(1 - \frac{b}{4ax^2} + \cdot\cdot\cdot\right) \\ &= 1 + \frac{1}{4x^2}\left(\frac{a}{b} - \frac{b}{a}\right) + \cdot\cdot\cdot\cdot . \end{aligned} \tag{203}$$

Thus we see that $z_0(x)$ varies between the value $\sqrt{a/b}$ at zero to the asymptotic value unity at infinity. The value (201) at $x = 1$ is thus the geometric mean. The slope at this point is found from differentiating (199) and is

$$\left(\frac{dz_0}{dx}\right)_{x=1} = -\frac{1}{2}\sqrt[4]{\frac{a}{b}} \cdot \left(\frac{a - b}{a + b}\right). \tag{204}$$

With this information it is a simple matter to plot $z_0(x)$ for any given case. This plot will depend only upon the ratio (a/b), and hence one such plot may cover a number of cases.

So far we have said nothing about the angle $\zeta(\omega)$ appearing in (197). This is evidently expressible as

$$\zeta(\omega) = \tfrac{1}{2}\{\measuredangle(R + jL\omega) + \measuredangle(G - jC\omega)\} \tag{205}$$

where the sign $\measuredangle$ is used to indicate the words "angle of." This expression may be put into the form

$$\begin{aligned} \zeta(\omega) &= \tfrac{1}{2}\measuredangle(R + jL\omega)(G - jC\omega) \\ &= \tfrac{1}{2}\measuredangle\{RG + LC\omega^2 + j\omega(LG - RC)\} \\ &= -\tfrac{1}{2}\tan^{-1}\left\{\frac{(RC - LG)\omega}{RG + LC\omega^2}\right\}, \end{aligned}$$

which, with the help of (148) and (149), may be written

$$\zeta(x) = -\tfrac{1}{2}\tan^{-1}\left(\frac{2nx}{1 + x^2}\right). \tag{206}$$

The most obvious characteristic of this function is the fact that

$$\zeta\left(\frac{1}{x}\right) = \zeta(x), \tag{207}$$

which will be useful in plotting. For small values of x we have

$$\zeta(x) \cong -\tfrac{1}{2} \tan^{-1} 2nx \cong -nx, \tag{207a}$$

from which we see that

$$\left.\begin{aligned} \zeta(0) &= 0 \\ \left(\frac{d\zeta}{dx}\right)_{x=0} &= -n \end{aligned}\right\}. \tag{208}$$

For large values of x, on the other hand,

$$\zeta(x) \cong -\tfrac{1}{2} \tan^{-1} \frac{2n}{x} \cong -\frac{n}{x}, \tag{207b}$$

which shows that

$$\zeta(\infty) = 0.$$

It is thus quite evident that $\zeta(x)$ has a maximum value somewhere between zero and infinity. This may be determined by the usual process of setting the derivative equal to zero. We find

$$\frac{d\zeta}{dx} = \frac{n(x^2 - 1)}{1 + 2(m^2 + n^2)x^2 + x^4} = 0,$$

from which the maximum occurs for

$$x_{\max} = \pm 1. \tag{209}$$

Substituting this back into (206) we have

$$\zeta_{\max} = \mp\tfrac{1}{2} \tan^{-1} n. \tag{210}$$

These results regarding the functions $z_0(x)$ and $\zeta(x)$ are summarized in the plot of Fig. 26 which is drawn for $m = 4$ and $a > b$. Here, as with the propagation function, we see that the ideal behavior will be obtained for the condition

$$a = b; \frac{R}{L} = \frac{G}{C},$$

for which obviously

$$\left.\begin{aligned} |Z_0| &= \sqrt{\frac{L}{C}} \\ \zeta &= 0 \end{aligned}\right\}, \tag{211}$$

i.e., the characteristic impedance reduces to a real constant, equal in value to that which obtains for the non-dissipative case.

For actual lines and cables the question arises as to how the param-

eters must be altered in order that the maximum variations in $|Z_0|$ and ζ may not exceed a prescribed value over a given frequency range. This problem may be solved in an analogous fashion to that used in solving the corresponding problem in connection with the propagation function. There we saw that for most practical purposes an approximate method would answer. Here we shall, therefore, discuss only the approximate solution.

A glance at Fig. 26 reveals the fact that with the characteristic impedance function the region of maximum distortion is in general also limited to the vicinity of the origin, just as with the attenuation and phase functions. At the higher frequencies both the magnitude and

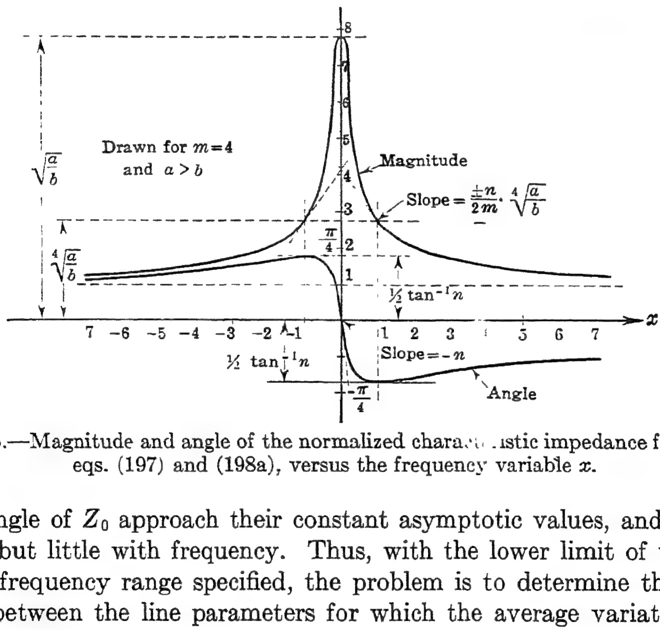


FIG 26.—Magnitude and angle of the normalized chara..., .istic impedance function, eqs. (197) and (198a), versus the frequency variable x.

the angle of Z_0 approach their constant asymptotic values, and hence vary but little with frequency. Thus, with the lower limit of the desired frequency range specified, the problem is to determine the relation between the line parameters for which the average variations in $z_0(x)$ and $\zeta(x)$ will stay within a prescribed tolerance.

An approximate solution to this problem is easily found if we make use of the relations (203) and (207b) for $z_0(x)$ and $\zeta(x)$, which are valid for large values of x. In the case of $z_0(x)$, the average variation over a frequency range from ω_1 to infinity is obviously half the total variation over this range. Denoting the average tolerance in $z_0(x)$ by δ_{z_0} we, therefore, have by (203)

$$\delta_{z_0} = \{z_0(\infty) - z_0(x_1)\}/\{z_0(\infty) + z_0(x_1)\}$$
$$\cong -\frac{1}{8x_1^2}\left(\frac{a}{b} - \frac{b}{a}\right), \tag{212}$$

from which, with the use of (148),

$$a^2 - b^2 = 8\omega_1^2\delta_{z_0}. \tag{212a}$$

When the lower frequency limit ω_1 and the maximum tolerance δ_{z_0} are specified, the amount by which the resistance must be reduced or the inductance increased may be found from

$$\left(\frac{R}{L}\right)_{\text{new}} = \sqrt{8\omega_1^2\delta_{z_0} + \frac{G^2}{C^2}}, \tag{212b}$$

where it is assumed that G and C are fixed.

In the case of $\zeta(x)$ the average variation between ω_1 and ∞ equals the average value throughout this range. If we denote this average value by δ_ζ, and use (207b) we have

$$\delta_\zeta = \tfrac{1}{2}\zeta(x_1) = -\frac{n}{2x_1}. \tag{213}$$

Since the algebraic sign is of no consequence if δ_ζ is to indicate a magnitude only, we get by using (148) and (149)

$$a - b = 4\omega_1\delta_\zeta. \tag{213a}$$

When ω_1, δ_ζ, and b are given we have

$$\left(\frac{R}{L}\right)_{\text{new}} = 4\omega_1\delta_\zeta + \frac{G}{C} \tag{213b}$$

for the determination of the required resistance to inductance ratio.

Since the average value δ_ζ will usually be small, the tangent of this angle may be replaced by the angle itself with sufficient accuracy as is also done in the relations (207a) and (207b). Hence to prescribe that δ_ζ shall be, say, 0.1 is equivalent to specifying that the average value of the imaginary part of the characteristic impedance shall not exceed 10 per cent of the magnitude of Z_0. The quantity $\delta_\zeta.100$, therefore, indicates the average value in percentage of $|Z_0|$ which the imaginary part of Z_0 equals throughout the specified frequency range.

8. Net effect of various tolerances. Having studied the variations in the propagation and characteristic impedance functions over specified frequency ranges, the reader will probably be interested in how these variations affect the amplitude and phase of the voltage and current ratios, i.e., in the net effect of the tolerances in α and Z_0 upon the transmission of voltage and current. In this section we shall attempt to show how this situation may be studied.

In section 17 of Chapter II we derived expressions for the ratios of input to output voltage and current, as well as for the ratios of the received voltage and current without the line to the corresponding

quantities with the line inserted between the sending- and receiving-end impedances. These ratios, which are referred to as *transmission* and *insertion* ratios, respectively, we repeat here for convenience, using the same equation numbers as before

$$\frac{E_S}{E_R} = e^{\alpha l} \cdot (1 + r_R)^{-1} \cdot (1 + r_R e^{-2\alpha l}), \tag{128a}$$

$$\frac{I_S}{I_R} = e^{\alpha l} \cdot (1 - r_R)^{-1} \cdot (1 - r_R e^{-2\alpha l}), \tag{131}$$

$$\frac{E_R'}{E_R} = \frac{I_R'}{I_R} = e^{\alpha l} \cdot (1 - r_S r_R)^{-1} \cdot (1 - r_S r_R e^{-2\alpha l}). \tag{130}$$

The volt-ampere ratios are readily formed from these.

The logarithms of these ratios we shall consider as representing the *actual* or *effective* propagation functions corresponding to either the transmission or insertion bases. These we denote by the letter γ, and distinguish in the following manner:

$$\left.\begin{aligned} \gamma' &= \ln\left(\frac{E_R'}{E_R}\right) = \ln\left(\frac{I_R'}{I_R}\right) = \ln\sqrt{\frac{E_R' I_R'}{E_R I_R}}; \\ \gamma_E &= \ln\left(\frac{E_S}{E_R}\right); \ \gamma_I = \ln\left(\frac{I_S}{I_R}\right); \ \gamma_{EI} = \ln\sqrt{\frac{E_S I_S}{E_R I_R}}. \end{aligned}\right\} \tag{214}$$

When the line is reasonably well terminated, the reflection coefficients are small. Approximate expressions may then be used to represent the logarithms of factors involving these coefficients if we apply the expansion

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$$

for the region of small values of x. Thus we find

$$\left.\begin{aligned} \gamma' &\cong \alpha l + r_S r_R(1 - e^{-2\alpha l}) \\ \gamma_E &\cong \alpha l - r_R(1 - e^{-2\alpha l}) \\ \gamma_I &\cong \alpha l + r_R(1 - e^{-2\alpha l}) \\ \gamma_{EI} &\cong \alpha l + \tfrac{1}{2} r_R^2(1 - e^{-4\alpha l}) \end{aligned}\right\}. \tag{214a}$$

For the first three of these only one term in the expansion is used; for the last, two terms are necessary since the odd terms contribute nothing. It should be noticed that in γ_E and γ_I the reflection is a *first order effect*, whereas in γ' and γ_{EI} it is a *second order effect*. This means that poor terminal conditions affect the quality of voltage and current transmission individually to a greater extent than they do either the volt-ampere transmission ratio or any of the insertion ratios.

The departure of the functions α and Z_0 from their ideal behavior gives rise to departures in the effective propagation function γ from its ideal behavior for the various cases summarized in (214a). If we denote the average values of the attenuation and phase functions throughout given frequency ranges by

$$\left.\begin{aligned} \alpha_{1\text{av}} &= \sqrt{RG} \cdot f_{1\text{av}} \\ \alpha_{2\text{av}} &= \omega\sqrt{LC} \cdot f_{2\text{av}} \end{aligned}\right\},$$

where f_{1av} and f_{2av} are the average values of the functions $f_1(x)$ and $f_2(x)$ discussed in section 3, then in the expression

$$\gamma = \alpha_{1\text{av}}l(1 + \delta_1) + j\alpha_{2\text{av}}l(1 + \delta_2) \tag{214b}$$

δ_1 and δ_2 represent the *net errors* (or decimal departures from the ideal) in the effective propagation function γ. The exact determination of δ_1 and δ_2 for the general case involves nothing fundamentally new but leads to expressions too cumbersome to reproduce here. We shall confine ourselves to an approximate treatment of the case for which the attenuation or line length are sufficiently great so that the factors $e^{-2\alpha l}$ or $e^{-4\alpha l}$ in (214a) may be neglected. The results of such a study may still be used as estimates in borderline cases.

For the propagation function α we have the representation

$$\alpha = \alpha_{1\text{av}}(1 + \delta_{\alpha 1}) + j\alpha_{2\text{av}}(1 + \delta_{\alpha 2}), \tag{215}$$

where $\delta_{\alpha 1}$ and $\delta_{\alpha 2}$ are the maximum decimal tolerances discussed in section 5. We recall that they are equal in magnitude but opposite in sign. Equation (180) gives $\delta_{\alpha 1}$ as

$$\delta_{\alpha 1} = -\delta_{\alpha 2} = \frac{(a - b)^2}{16\,\omega_1^2},$$

where ω_1 is the lowest essential frequency.

In order to get an average maximum value for the reflection coefficients, we write for the characteristic impedance

$$\begin{aligned} Z_0 &= |Z_0|_{\text{av}} \cdot (1 + \delta_{z_0}) \cdot \underline{|\delta_\zeta} \\ &\cong |Z_0|_{\text{av}} \cdot (1 + \delta_{z_0})(1 + j\delta_\zeta) \\ &\cong |Z_0|_{\text{av}} \cdot (1 + \delta_{z_0} + j\delta_\zeta) \end{aligned}$$

where $|Z_0|_{\text{av}}$ is its average magnitude in the prescribed range. If the terminal impedance Z_R is made equal to this average value, the reflection coefficient becomes

$$r_R \cong -\frac{\delta_{z_0} + j\delta_\zeta}{2 + \delta_{z_0} + j\delta_\zeta} \cong -\tfrac{1}{2}(\delta_{z_0} + j\delta_\zeta). \tag{215a}$$

Assuming that the line is similarly terminated at the sending end, the coefficient r_S has the same value. Thus, substituting (215) and (215a) into (214a), and dropping the exponential terms, the net errors according to (214b) are for the function γ'

$$\delta_1 = \delta_{\alpha 1} + \frac{\delta_{z_0}{}^2 - \delta_\zeta{}^2}{4\alpha_{1\,\mathrm{av}} l}; \quad \delta_2 = \delta_{\alpha 2} + \frac{\delta_{z_0}\delta_\zeta}{2\alpha_{2\,\mathrm{av}} l}; \tag{216}$$

for the functions γ_E and γ_I (top and bottom signs respectively)

$$\delta_1 = \delta_{\alpha 1} \pm \frac{\delta_{z_0}}{2\alpha_{1\,\mathrm{av}} l}; \quad \delta_2 = \delta_{\alpha 2} \pm \frac{\delta_\zeta}{2\alpha_{2\,\mathrm{av}} l}; \tag{216a}$$

and for the function γ_{EI}

$$\delta_1 = \delta_{\alpha 1} + \frac{\delta_{z_0}{}^2 - \delta_\zeta{}^2}{8\alpha_{1\,\mathrm{av}} l}; \quad \delta_2 = \delta_{\alpha 2} + \frac{\delta_{z_0}\delta_\zeta}{4\alpha_{2\,\mathrm{av}} l}. \tag{216b}$$

These results show that for long lines and high attenuation the net errors are substantially given by $\delta_{\alpha 1}$ and $\delta_{\alpha 2}$ in every case. In order to evaluate them still further, we shall assume that a is large compared to b, which is justified even though the line parameters are adjusted for reasonably small values of $\delta_{\alpha 1}$ and $\delta_{\alpha 2}$, and then have

$$\left.\begin{aligned} \delta_{\alpha 1} &= -\delta_{\alpha 2} = \frac{a^2}{16\omega_1{}^2} \\ \delta_{z_0} &= -\frac{a^2}{8\omega_1{}^2} = -2\delta_{\alpha 1} \\ \delta_\zeta &= -\delta_{\alpha 1}{}^{1/2}\left(1 - \frac{16}{3}\delta_{\alpha 1}\right) \end{aligned}\right\},$$

where (212) is used for δ_{z_0}, and an additional term in (207b) is calculated for the determination of δ_ζ according to (213). This is necessary in order to evaluate (216) and (216b) which involve the difference of $\delta_{z_0}{}^2$ and $\delta_\zeta{}^2$. The second term in δ_ζ, therefore, contributes a correction of the same order of magnitude as $\delta_{z_0}{}^2$.

For $\alpha_{1\,\mathrm{av}}$ and $\alpha_{2\,\mathrm{av}}$ we can use

$$\alpha_{1\,\mathrm{av}} \cong \sqrt{RG}\, m \cong \frac{R}{2}\sqrt{\frac{C}{L}}$$

$$\alpha_{2\,\mathrm{av}} \cong \omega_1\sqrt{LC} = \tfrac{1}{2}\,\delta_{\alpha 1}{}^{-1/2}\left(\frac{R}{2}\sqrt{\frac{C}{L}}\right)$$

where the frequency limit ω_1 is substituted in order to obtain the maximum error. With these values we find for γ' the errors

$$\delta_1 = \delta_{\alpha 1}\left\{1 + \left(\frac{11}{3}\delta_{\alpha 1} - \frac{1}{4}\right)\left(\frac{R}{2}\sqrt{\frac{C}{L}}l\right)^{-1}\right\}; \quad \delta_2 = \delta_{\alpha 2}\left\{1 + 2\delta_{\alpha 2}\left(\frac{R}{2}\sqrt{\frac{C}{L}}l\right)^{-1}\right\}, \tag{216c}$$

for γ_E and γ_I

$$\delta_1 = \delta_{\alpha 1}\left\{1 \mp \left(\frac{R}{2}\sqrt{\frac{C}{L}}l\right)^{-1}\right\};\ \delta_2 = \delta_{\alpha 2}\left\{1 \pm \left(\frac{R}{2}\sqrt{\frac{C}{L}}l\right)^{-1}\right\}. \quad (216d)$$

and for γ_{EI}

$$\delta_1 = \delta_{\alpha 1}\left\{1 + \left(\frac{11}{6}\delta_{\alpha 1} - \frac{1}{8}\right)\left(\frac{R}{2}\sqrt{\frac{C}{L}}l\right)^{-1}\right\};\ \delta_2 = \delta_{\alpha 2}\left\{1 + \delta_{\alpha 2}\left(\frac{R}{2}\sqrt{\frac{C}{L}}l\right)^{-1}\right\}. \quad (216e)$$

As an illustration of this result consider the numerical example given at the end of section 5, page 98, where the necessary R/L ratio to make $\delta_\alpha = 0.05$ for $\omega_1 = 2\pi \cdot 100$ was calculated for the open line data on page 90. Assuming that the resistance remains unchanged, this calls for a new inductance parameter

$$L_{\text{new}} = 3.78 \times 0.004 = 0.01512.$$

With this value

$$\frac{R}{2}\sqrt{\frac{C}{L}} = 5\sqrt{\frac{0.008}{0.01512}} \cdot 10^{-3} = 0.00364.$$

This line will have to be several hundred miles long before the reflection effects due to imperfect terminations become negligible. Suppose the line is 275 miles long. Then

$$e^{-2\alpha_1 l} = 0.135;\ e^{-4\alpha_1 l} = 0.018$$

so that there is little error in dropping the exponential terms in (214a). Using $\delta_{\alpha 1} = -\delta_{\alpha 2} = 0.05$ we have for γ'

$$\delta_1 = 0.05\ (1 - 0.23) = 3.8\%$$
$$\delta_2 = -0.05\ (1 - 0.1) = -4.5\%$$

for γ_E

$$\delta_1 = 0.05\ (1 - 1) = 0\%$$
$$\delta_2 = -0.05\ (1 + 1) = 10\%$$

for γ_I

$$\delta_1 = 0.05\ (1 + 1) = 10\%$$
$$\delta_2 = -0.05\ (1 - 1) = 0\%$$

and for γ_{EI}

$$\delta_1 = 0.05\ (1 - 0.11) = 4.5\%$$
$$\delta_2 = -0.05\ (1 - 0.05) = 4.8\%$$

These are the approximate net amounts of amplitude and phase distortion over this circuit at the lowest frequency. At the highest frequency the tolerances $\delta_{\alpha 1}$ and $\delta_{\alpha 2}$ have opposite signs while δ_ζ is practically zero. The errors in γ' and γ_{EI} will not change appreciably. Those in γ_E and γ_I will change by about the same amount owing to a second order

effect. It is interesting to note that sometimes the reflection effects tend to cancel the errors due to the propagation function.

From this discussion we see that reflections have little effect upon the quality of transmission only when the line is long or the attenuation high. It may, of course, be that in a given situation the degree of mismatch at the line terminals is an even more important factor than the quality of transmission. This will be true at any terminal where line balance is an important item, as for example at repeater points in a two-wire single channel carrier system. There the repeater gain is definitely limited by the accuracy of balance which can be obtained between the characteristic impedance of the line and the balancing network. The situation can obviously be met by either improving the characteristics of the line sufficiently or by designing a more elaborate line balancing network. The latter process, the theory of which we shall discuss in a later chapter, is the one favored in recent practice.

9. Approximations in the consideration of the attenuation and phase functions. It oftens occurs that the line parameters in an actual case have such values that within certain frequency ranges one or more may be neglected without having the resulting approximate expressions for α_1 and α_2 be in error by more than a tolerable amount. When this is so, the treatment of the transmission problem may be considerably simplified numerically. It is, therefore, of value to establish criteria whereby we shall be able to predetermine whether or not such approximations are possible when the essential frequency range and the maximum allowable error are specified. This we propose to do now.

Consider the following analytic forms for α_1 and α_2 which follow from (146b), (147b), (150a), and (151a) with the use of (148), namely

$$\left.\begin{aligned} \alpha_1 &= \sqrt{RGx}\,(\sqrt{m^2+y^2}-y)^{1/2} \\ \alpha_2 &= \sqrt{RGx}\,(\sqrt{m^2+y^2}+y)^{1/2} \end{aligned}\right\}. \tag{217}$$

According to (153) the frequency region for which x lies in the vicinity of unity is that for which y lies in the vicinity of zero. For this vicinity the right-hand factors in (217) may be expanded as follows

$$(\sqrt{m^2+y^2}\pm y)^{1/2} = \sqrt{m}\left\{1 \pm \frac{1}{2}\left(\frac{y}{m}\right) + \frac{1}{8}\left(\frac{y}{m}\right)^2 \pm \cdots\right\}. \tag{218}$$

Then if we recognize that

$$\begin{aligned} \sqrt{RGmx} &= \sqrt{\frac{RC\omega}{2}}\left(1+\frac{b}{a}\right)^{1/2} \\ &= \sqrt{\frac{RC\omega}{2}}\left\{1+\frac{1}{2}\left(\frac{b}{a}\right)-\frac{1}{8}\left(\frac{b}{a}\right)^2+\cdots\right\}, \end{aligned} \tag{219}$$

and assume

$$\frac{b}{a} << 1 \tag{220}$$

so that to a first approximation the series in (218) and (219) may be broken off after the second terms, we can write

$$\left.\begin{aligned} \alpha_1 &\cong \sqrt{\frac{RC\omega}{2}}\left\{1 + \frac{1}{2}\left(\frac{b}{a} - \frac{y}{m}\right)\right\} \\ \alpha_2 &\cong \sqrt{\frac{RC\omega}{2}}\left\{1 + \frac{1}{2}\left(\frac{b}{a} + \frac{y}{m}\right)\right\} \end{aligned}\right\}. \tag{221}$$

But, using (148) and (153) again, we see that

$$\frac{y}{m} = \left(\frac{\omega}{a} - \frac{b}{\omega}\right)\left(1 + \frac{b}{a}\right)^{-1}, \tag{222}$$

and expanding the right-hand factor in this we have

$$\left(\frac{b}{a} \pm \frac{y}{m}\right) = \left(\frac{b}{a}\right) \pm \left(\frac{\omega}{a} - \frac{b}{\omega}\right) \mp \left(\frac{b}{a}\right)\left(\frac{\omega}{a} - \frac{b}{\omega}\right) \pm \cdots .$$

Retaining the first two terms in this expansion, we may write

$$\left.\begin{aligned} \alpha_1 &= \sqrt{\frac{RC\omega}{2}}(1 + \delta_1) \\ \alpha_2 &= \sqrt{\frac{RC\omega}{2}}(1 + \delta_2) \end{aligned}\right\}, \tag{223}$$

and for the decimal errors δ_1 and δ_2 have

$$\left.\begin{aligned} \delta_1 &\cong \frac{1}{2}\left(\frac{b}{a} - \frac{\omega}{a} + \frac{b}{\omega}\right) \\ \delta_2 &\cong \frac{1}{2}\left(\frac{b}{a} + \frac{\omega}{a} - \frac{b}{\omega}\right) \end{aligned}\right\}. \tag{224}$$

These results state that, if $(b/a) \ll 1$, we may use for α_1 and α_2 the approximate expression

$$\alpha_1 = \alpha_2 = \sqrt{\frac{RC\omega}{2}} \tag{225}$$

where the errors and their dependence upon frequency are approximately expressed by (224). Recalling the original form (142) for α, we recognize that the form (225) is the result of neglecting the line parameters L and G. In practice this is frequently done in the treatment of cable problems, where it is pointed out that, owing to the close proximity of the conductors and the relatively good insulation, the inductance

and leakage are so small as to be altogether negligible in comparison with the resistance and capacitance parameters. This may very well be so, but it is nevertheless essential to demonstrate the correctness of such a procedure, particularly in doubtful cases. Furthermore it is all-important to note that whenever we use the term "negligible" it is implied that, although an error is involved, it is for the case at hand so small that it does not affect the result to a greater degree than can be tolerated. Thus the question as to whether or not certain parameters may be neglected is not adequately met unless we are able to specify the maximum magnitude of the error which will be involved when this approximation is made over a specified frequency range. The latter is a particularly significant factor in this connection, for although it may be quite adequate to neglect L and G in a cable over a certain frequency range, the same may not be true for another range unless the maximum tolerance is also modified.

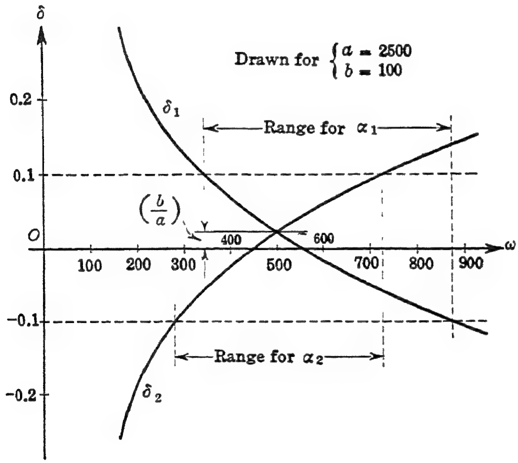


FIG. 27.—Errors in the attenuation (δ_1) and phase (δ_2) functions due to neglecting the parameters L and G showing their variations with frequency. Drawn for the data (230).

This situation is expressed by the relations (224), the significance of which is most easily understood by referring to Fig. 27 where these errors are plotted versus frequency for a numerical example. From these curves we see that the errors are zero at one frequency, and increase in magnitude as this frequency is departed from. Hence, if we specify a given maximum error (in the figure this is chosen as $\delta = \pm 0.1$) we will find corresponding frequency ranges within which this error is not exceeded. We note from the figure that these ranges are not quite the same for α_1 and α_2. This is due to the term (b/a) in (224). The smaller (b/a), the more nearly identical will the ranges for α_1 and α_2 be. It is also interesting to note that the error has opposite signs at the two frequency limits.

When the maximum tolerance is specified and it is desired to determine the corresponding frequency ranges, we must solve the relations

(224) for ω. This gives for the ranges for α_1 and α_2 respectively

$$\frac{-(2a\delta_1 - b) + \sqrt{(2a\delta_1 - b)^2 + 4ab}}{2} \leq \omega \leq \frac{(2a\delta_1 + b) + \sqrt{(2a\delta_1 + b)^2 + 4ab}}{2} \quad (226)$$

and

$$\frac{-(2a\delta_2 + b) + \sqrt{(2a\delta_2 + b)^2 + 4ab}}{2} \leq \omega \leq \frac{(2a\delta_2 - b) + \sqrt{(2a\delta_2 - b)^2 + 4ab}}{2} \quad (227)$$

where the δ's represent the magnitudes of the errors only. If

$$b \ll 2a\delta \quad (228)$$

these expressions may be approximated by neglecting b in comparison with $2a\delta$, giving the following range common to α_1 and α_2

$$a\delta\left\{\sqrt{1 + \frac{b}{a\delta^2}} - 1\right\} \leq \omega \leq a\delta\left\{\sqrt{1 + \frac{b}{a\delta^2}} + 1\right\}. \quad (229)$$

This may be further simplified if

$$b \ll a\delta^2 \quad (228a)$$

in which case the radicals may be expanded and only the first two terms retained. Then the common range is expressed by

$$\frac{b}{2\delta}\left(1 - \frac{b}{4a\delta^2}\right) \leq \omega \leq 2a\delta\left(1 + \frac{b}{4a\delta^2}\right). \quad (229a)$$

Note that where the use of (229) or (229a) is justified the product of the limiting frequencies equals ab; i.e., the point $x = 1$ becomes the geometric mean between the limits of the range. This shows that the vicinity of $x = 1$ is that region within which the inductance and leakage are negligible provided to begin with

$$\left(\frac{b}{a}\right) = \frac{LG}{RC} \ll 1.$$

In order to illustrate these results consider first the data for an open telephone line

$$\left.\begin{array}{l} R = 10;\ L = 0.004 \\ G = 0.8 \cdot 10^{-6};\ C = 0.008 \cdot 10^{-6} \end{array}\right\}. \quad (230)$$

Here

$$a = \frac{R}{L} = 2500;\ b = \frac{G}{C} = 100;\ \frac{b}{a} = \frac{1}{25}.$$

Suppose we let

$$\delta_1 = \delta_2 = 0.1,$$

i.e., fix the maximum error at 10 per cent. Although

$$\frac{b}{2a\delta} = \frac{1}{5}$$

we shall not consider (228) sufficiently justified. Hence we use the forms (226) and (227) and find for α_1 the range

$$\left.\begin{array}{rl} & 339 \leq \omega \leq 883 \\ \text{or} & 54 \leq f \leq 140 \end{array}\right\}, \tag{231}$$

and for α_2 the range

$$\left.\begin{array}{rl} & 283 \leq \omega \leq 739 \\ \text{or} & 45 \leq f \leq 118 \end{array}\right\}. \tag{232}$$

These results will be seen to check the Fig. 27 which is drawn for this case. The frequency ranges thus found are not very extensive, to be sure, but it is rather surprising that both the inductance and the leakage are at all negligible on an ordinary open telephone line, even to within a maximum error of 10 per cent and over a limited frequency range.

In contrast to this, consider the following telephone cable data

$$\left.\begin{array}{l} R = 171;\ L = 0.001 \\ G = 1.75 \cdot 10^{-6};\ C = 0.073 \cdot 10^{-6} \end{array}\right\}. \tag{233}$$

Then

$$a = \frac{R}{L} = 171{,}000;\ b = \frac{G}{C} = 24;\ \frac{b}{a} = \frac{1}{7125}.$$

Again let

$$\delta_1 = \delta_2 = 0.1.$$

Here

$$\frac{b}{a\delta^2} = \frac{1}{71},$$

so that the condition (228a) is sufficiently satisfied, and hence we can use the relation (229a) for the determination of the common frequency range. We find

$$\begin{array}{rl} & 140 \leq \omega \leq 34{,}200 \\ \text{or} & 22 \leq f \leq 5450. \end{array} \tag{234}$$

Thus we see that on this cable circuit the inductance and leakage are negligible, so far as the propagation function is concerned, to within a maximum error of 10 per cent over a frequency range of 22 to 5450

cycles per second. This is a considerable range, and adequately covers the voice spectrum, but it is nevertheless significant that below 22 and above 5450 cycles per second even this cable circuit cannot be treated by considering its resistance and capacitance alone without introducing an error in excess of 10 per cent.

If the conditions or the requirements are such that for a given circuit the above approximations cannot be made, then it may still be possible that for this circuit the leakage parameter alone is negligible. In order to study this possibility, consider again the forms (217), but in the slightly modified form

$$\left.\begin{aligned} \alpha_1 &= \sqrt{RGmx}\left(\sqrt{1+\left(\frac{y}{m}\right)^2} - \frac{y}{m}\right)^{1/2} \\ \alpha_2 &= \sqrt{RGmx}\left(\sqrt{1+\left(\frac{y}{m}\right)^2} + \frac{y}{m}\right)^{1/2} \end{aligned}\right\}. \tag{235}$$

Now let

$$\frac{y}{m} = y_a + y_b$$

and

$$\left.\begin{aligned} \frac{y}{m} &= y_a + y_b \\ y_a &= \frac{\omega}{a} \end{aligned}\right\}, \tag{236}$$

so that by (222)

$$y_b = -\left(\frac{b}{\omega} + \frac{b}{a}\frac{\omega}{a}\right)\left(1 + \frac{b}{a}\right)^{-1}. \tag{237}$$

Since by using (148) we may modify (219) so as to give either

$$\sqrt{RGmx} = \frac{R}{2}\sqrt{\frac{C}{L}}(2y_a)^{1/2}\left(1+\frac{b}{a}\right)^{1/2} \tag{238}$$

or

$$\sqrt{RGmx} = \omega\sqrt{LC}\,(2y_a)^{-1/2}\left(1+\frac{b}{a}\right)^{1/2}, \tag{239}$$

we recognize that if we can determine the frequency range for which y_b is negligible, we will also have the region of negligible G. This is so because y_b vanishes for $b = 0$, which is the same as $G = 0$. For this condition (238) and (239) merely drop their last factor.

In order to determine this range and the corresponding errors in α_1 and α_2, we assume

$$y_b \ll y_a$$

and expand the second factors in (235) about the point $y_b = 0$, retaining only the first two terms. This gives

$$\left(\sqrt{1+\left(\frac{y}{m}\right)^2} \pm \frac{y}{m}\right)^{1/2} = (\sqrt{1+y_a^2} \pm y_a)^{1/2}\left(1 \pm \frac{y_b}{2\sqrt{1+y_a^2}}\right). \tag{240}$$

Substituting this together with (237), (238), and (239) into (235), and using the first two terms of

$$\left(1+\frac{b}{a}\right)^{1/2} = 1 + \frac{b}{2a} - \frac{1}{8}\left(\frac{b}{a}\right)^2 + \cdots,$$

we find

$$\left.\begin{aligned} \alpha_1 &= \frac{R}{2}\sqrt{\frac{C}{L}}\,(2y_a)^{1/2}(\sqrt{1+y_a^2} - y_a)^{1/2}(1+\delta_1) \\ \text{and}\qquad \alpha_2 &= \omega\sqrt{LC}(2y_a)^{-1/2}(\sqrt{1+y_a^2} + y_a)^{1/2}(1+\delta_2) \end{aligned}\right\} \tag{241}$$

with

$$\delta_{1,2} = \frac{b}{2a} \pm \frac{\left(\dfrac{b}{\omega} + \dfrac{b}{a}\dfrac{\omega}{a}\right)}{2\left(1+\dfrac{b}{a}\right)\sqrt{1+\left(\dfrac{\omega}{a}\right)^2}}. \tag{242}$$

The forms (241) are the result of neglecting G, and (242) determines the errors and their frequency dependence. A study of (242) shows that the errors are infinite at zero frequency, and decrease uniformly with increasing frequency approaching an asymptotic value as the frequency tends toward infinity. More specifically

$$\left.\begin{aligned} &\text{for } \omega \to \infty \\ &\delta_1 \to \frac{b}{a} \\ &\delta_2 \to \frac{1}{2}\left(\frac{b}{a}\right)^2 \end{aligned}\right\}, \tag{243}$$

and to within a sufficient approximation

$$\left.\begin{aligned} &\text{for } \omega \to 0 \\ &\delta_1 \to \left(\frac{b}{2a} + \frac{b}{2\omega}\right) \\ &\delta_2 \to \left(\frac{b}{2a} - \frac{b}{2\omega}\right) \end{aligned}\right\}. \tag{244}$$

Since it is assumed to start with that $b \ll a$, the errors (243) are negligible. Hence if the asymptotic values (243) are sufficiently small for the case at hand, the approximate relations (244) may be used to determine the lower frequency limit below which a prescribed error will be exceeded. This is easily done by solving (244) for ω. Thus the range for negligible G is given by

$$\left.\begin{aligned} \frac{ab}{2a\delta_1 - b} &\leq \omega \leq \infty \\ \frac{ab}{2a\delta_2 + b} &\leq \omega \leq \infty \end{aligned}\right\}, \tag{245}$$

or if

$$\frac{b}{2a\delta} << 1, \tag{246}$$

these are with sufficient accuracy

$$\left.\begin{aligned} \frac{b}{2\delta_1}\left(1+\frac{b}{2a\delta_1}\right) \leq \omega \leq \infty \\ \frac{b}{2\delta_2}\left(1-\frac{b}{2a\delta_2}\right) \leq \omega \leq \infty \end{aligned}\right\}. \tag{245a}$$

Applying these results to the data (230) for an average open line we find that the leakage parameter is negligible to within a maximum error of 10 per cent in α_1 for

$$\left.\begin{aligned} 600 \leq \omega \leq \infty \\ 95 \leq f \leq \infty \end{aligned}\right\},$$

and in α_2 for

$$\left.\begin{aligned} 400 \leq \omega \leq \infty \\ 64 \leq f \leq \infty \end{aligned}\right\}.$$

This is quite remarkable for an ordinary open telephone line. Of course, as pointed out in Chapter I, the leakage parameter increases considerably with frequency. The value in (230) is for a frequency of about 800 cycles. Hence, although the variation is not considered in the above calculations, it is safe to say that over the voice range, and perhaps considerably beyond, the leakage parameter is still negligible. When the variation with frequency is known, the various forms discussed above may obviously still be applied by using a trial method. Thus the considerable variation of the leakage parameter with frequency is found for many practical cases to cause no annoyance whatsoever.

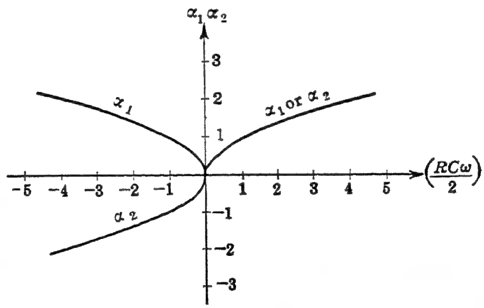


FIG. 28.—Attenuation and phase as functions of frequency when L and G are neglected and R and C assumed constant.

This situation is even more true in a cable. Using the data (233) we find by (245a) that the leakage parameter is negligible in both α_1 and α_2 to within a maximum error of 10 per cent for all frequencies above 22 cycles per second.

Thus we see that, for certain frequency ranges and tolerances, either G alone or both L and G may be neglected in the analysis of a given facility so far as the propagation function is concerned. When this is so the propagation function and consequently the numerical work will be simplified. For example, when both L and G are negligible, we have the one simple function (225) as a representation for both α_1 and α_2. A plot of this is shown in Fig. 28. It is simply a rectangular parabola tipped over on its side. Note that for negative frequencies α_1 is represented by the upper, while α_2 is represented by the lower branch of the parabola. This must be so because α_1 is an even and α_2 an odd function.

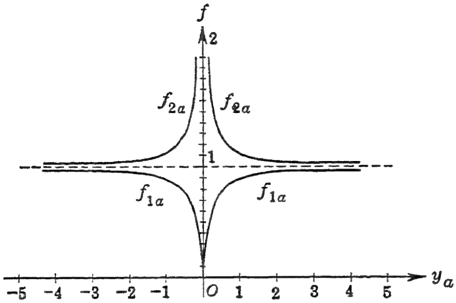


FIG. 29.—Departure functions for the attenuation and phase versus the frequency variable y_a, eq. (236), when the line parameter G is neglected.

When only G is negligible the simplification is not so great. We then have the expressions (241) with the δ's dropped. If we write

$$\left.\begin{aligned} \alpha_1 &= \frac{R}{2}\sqrt{\frac{C}{L}} \cdot f_{1a} \\ \alpha_2 &= \omega\sqrt{LC} \cdot f_{2a} \end{aligned}\right\}, \tag{247}$$

then

$$\left.\begin{aligned} f_{1a} &= (2y_a)^{1/2}(\sqrt{1 + y_a^2} - y_a)^{1/2} \\ f_{2a} &= (2y_a)^{-1/2}(\sqrt{1 + y_a^2} + y_a)^{1/2} \end{aligned}\right\}. \tag{248}$$

Plots of f_{1a} and f_{2a} are shown in Fig. 29. Note that

$$f_{1a} f_{2a} = 1, \tag{249}$$

i.e., the functions are reciprocals of each other. This simplifies the plotting considerably. It should be pointed out here that the functions (248) are not even functions of y_a. In fact for negative values of y_a they become imaginary. This rather puzzling point is cleared up if we consider (247) and form $\alpha_1 + j\alpha_2$. Then for a negative y_a the first term becomes imaginary and the second real, i.e., α_1 and α_2 interchange places but their analytic forms are left unchanged. The functions in Fig. 29 for negative y_a-values are those which are obtained after this con-

version process has been carried out. They are, of course, images of the corresponding functions for positive arguments.

10. Approximations in the consideration of the characteristic impedance function. When the inductance and leakage parameters are negligible in the determination of the characteristic impedance function, then a simplification in its analytic form is also possible. The maximum error and corresponding frequency range will, however, be different from those for the propagation function.

Considering the magnitude function (199), this may be put into the form

$$z_0 = \left(\frac{a}{\omega}\right)^{1/2}\left(1 + \frac{\omega^2}{a^2}\right)^{1/4}\left(1 + \frac{b^2}{\omega^2}\right)^{-1/4}, \tag{250}$$

so that by (198a) we have for the magnitude of the characteristic impedance

$$|Z_0| = \sqrt{\frac{R}{C\omega}}\left(1 + \frac{\omega^2}{a^2}\right)^{1/4}\left(1 + \frac{b^2}{\omega^2}\right)^{-1/4}. \tag{251}$$

Using the first two terms of the expansions

$$\left.\begin{aligned}\left(1 + \frac{\omega^2}{a^2}\right)^{1/4} &= 1 + \frac{1}{4}\frac{\omega^2}{a^2} + \cdots \\ \left(1 + \frac{b^2}{\omega^2}\right)^{-1/4} &= 1 - \frac{1}{4}\frac{b^2}{\omega^2} + \cdots\end{aligned}\right\} \tag{252}$$

and writing

$$|Z_0| = \sqrt{\frac{R}{C\omega}}\,(1 + \delta), \tag{251a}$$

we have

$$\delta = \frac{1}{4}\left(\frac{\omega^2}{a^2} - \frac{b^2}{\omega^2}\right) \tag{253}$$

for the error and its dependence upon frequency.

The form (251a) without the factor $(1 + \delta)$ is obviously the resulting expression for the magnitude of the characteristic impedance for $L = G = 0$. From (253) we find that the frequency region within which this approximation is valid to within a maximum decimal error δ is defined by

$$\sqrt{(4a^4\delta^2 + a^2b^2)^{1/2} - 2a^2\delta} \leq \omega \leq \sqrt{(4a^4\delta^2 + a^2b^2)^{1/2} + 2a^2\delta}. \tag{254}$$

When $b \ll 2a\delta$, this range is approximately given by

$$\frac{b}{2\sqrt{\delta}}\left(1 - \frac{b^2}{32a^2\delta^2}\right) \leq \omega \leq 2a\sqrt{\delta}\left(1 + \frac{b^2}{32a^2\delta^2}\right). \tag{254a}$$

Note that the product of these frequency limits equals ab. Thus the point $x = 1$ ($\omega = \sqrt{ab}$) is the geometric mean frequency of this region.

Using the data (230) for the open line and letting $\delta = 0.1$, we find that L and G are negligible in $|Z_0|$ with a maximum error of 10 per cent for the range

$$\left.\begin{aligned} 158 \leq \omega \leq 1580 \\ 25 \leq f \leq 250 \end{aligned}\right\}. \tag{255}$$

For the cable data (233), on the other hand, we find for the same maximum error the range

$$\left.\begin{aligned} 38 \leq \omega \leq 108{,}000 \\ 6 \leq f \leq 17{,}200 \end{aligned}\right\}. \tag{256}$$

Let us now see whether L and G may under certain circumstances be neglected in the formulation of the angle of the characteristic impedance. From (205) we may easily get

$$\zeta(\omega) = \frac{1}{2}\left\{\tan^{-1}\left(\frac{\omega}{a}\right) - \tan^{-1}\left(\frac{\omega}{b}\right)\right\}. \tag{257}$$

Assuming

$$\left(\frac{\omega}{a}\right) \ll 1, \text{ and } \left(\frac{\omega}{b}\right) \gg 1,$$

we may expand the inverse tangent functions and get

$$\left.\begin{aligned} \tan^{-1}\left(\frac{\omega}{a}\right) &= \left(\frac{\omega}{a}\right) - \frac{1}{3}\left(\frac{\omega}{a}\right)^3 + \cdots \\ \tan^{-1}\left(\frac{\omega}{b}\right) &= \frac{\pi}{2} - \left(\frac{\omega}{b}\right)^{-1} + \frac{1}{3}\left(\frac{\omega}{b}\right)^{-3} - \cdots \end{aligned}\right\}. \tag{258}$$

If we retain only the linear terms we have for (257)

$$\zeta(\omega) = -\frac{\pi}{4} + \frac{1}{2}\left(\frac{\omega}{a} + \frac{b}{\omega}\right),$$

or if we write

$$\zeta(\omega) = -\frac{\pi}{4}(1 + \delta) \tag{259}$$

we get

$$\delta = -\frac{2}{\pi}\left(\frac{\omega}{a} + \frac{b}{\omega}\right). \tag{260}$$

The value $-\frac{\pi}{4}$ in (259) we recognize as the angle of the characteristic impedance function for neglected L and G. The relation (260) gives the error and its dependence upon frequency. Solving for the frequency range we find

$$\frac{1}{4}(\pi a\delta - \sqrt{\pi^2 a^2 \delta^2 - 16ab}) \leq \omega \leq \frac{1}{4}(\pi a\delta + \sqrt{\pi^2 a^2 \delta^2 - 16ab}), \tag{261}$$

or if $b \ll a\delta^2$ this reduces to the following approximate form

$$\frac{2b}{\pi\delta}\left(1 + \frac{4b}{\pi^2 a\delta^2}\right) \leq \omega \leq \frac{\pi a\delta}{2}\left(1 - \frac{4b}{\pi^2 a\delta^2}\right). \tag{261a}$$

Note that (261) has solutions only when

$$\left(\frac{b}{a}\right) \leq \left(\frac{\pi\delta}{4}\right)^2, \tag{262}$$

i.e., unless the ratio b/a is smaller than this value, no range exists within which L and G are negligible for the angle ζ.

To illustrate this, consider again the open line data (230), and assume a maximum error of 10 per cent. Here

$$\frac{b}{a} = 0.04$$

and

$$\left(\frac{\pi\delta}{4}\right)^2 = 0.00617.$$

Hence the criterion (262) is not satisfied, and there exists no range for which L and G may be neglected in ζ to within a maximum error of 10 per cent. Note, however, that for the functions α_1, α_2, and $|Z_0|$ we did find ranges within which this same error was not exceeded! The student should thus clearly recognize that, although we may be able in a given instance to neglect L and G in some functions, it does not follow that this may be generally done. This emphasizes again the point made earlier that when we say that certain parameters are negligible it is absolutely essential to mention the function for which these are neglected as well as the maximum error which we are willing to tolerate and the corresponding frequency range.

For the cable data (233) we have with respect to the angle ζ and a maximum error of 10 per cent.

$$\left(\frac{b}{a}\right) = 0.00014$$

$$\left(\frac{\pi\delta}{4}\right)^2 = 0.00617.$$

Here the criterion (262) is obviously satisfied. Also the approximate form (261a) is applicable. We thus find the range

$$\left.\begin{aligned} 153 \leq \omega &\leq 26{,}850 \\ 24 \leq f &\leq 4{,}275 \end{aligned}\right\} \tag{263}$$

within which both L and G are negligible in ζ with a maximum error of 10 per cent for the cable whose data are given by (233).

Finally we consider the range for which G alone may be negligible in either the magnitude or angle of the characteristic impedance function. For the magnitude we see from (250) that the leakage parameter will be negligible (this means that b is negligible) when the last factor may be replaced by unity. Using for this factor the first two terms of the expansion in (252) and writing for brevity

$$\left(\frac{\omega}{a}\right) = y_a$$

as has already been done in (236), we have

$$|Z_0| = \sqrt{\frac{L}{C}} \cdot \sqrt[4]{\frac{1 + y_a^2}{y_a^2}}\,(1 + \delta) \tag{264}$$

with

$$\delta = -\frac{b^2}{4\omega^2}. \tag{265}$$

This error vanishes at high frequencies. The frequency range for a given error, without regard to the sign of this error, therefore is

$$\frac{b}{2\sqrt{\delta}} \leq \omega \leq \infty. \tag{265a}$$

Note that this is practically the same as the lower frequency limit in the region for negligible L and G for $|Z_0|$ as expressed by (254a). Hence for the open line data (230), and with $\delta = 0.1$, we have from (255) that G alone is negligible in $|Z_0|$ for

$$\left.\begin{array}{r} 158 \leq \omega \leq \infty \\ 25 \leq f \leq \infty \end{array}\right\}, \tag{266}$$

and for the cable data (233) we get from (256) the range

$$\left.\begin{array}{r} 38 \leq \omega \leq \infty \\ 6 \leq f \leq \infty \end{array}\right\}. \tag{267}$$

For a possible range of negligible G in the angle ζ, consider again the form (257) and expand only the second anti-tangent according to (258), retaining the first two terms. This gives

$$\zeta \cong \frac{1}{2}\left(\tan^{-1}\left(\frac{\omega}{a}\right) - \frac{\pi}{2} + \frac{b}{\omega}\right),$$

or if we write

$$\zeta = \left(-\frac{\pi}{4} + \frac{1}{2}\tan^{-1}\frac{\omega}{a}\right)(1 + \delta) \tag{268}$$

we have

$$\delta = \frac{b}{\omega\left(\tan^{-1}\frac{\omega}{a} - \frac{\pi}{2}\right)}. \tag{269}$$

This error is shown plotted versus ω/a in Fig. 30, from which we see that with increasing frequency the asymptotic value b/a is approached, while for smaller frequencies the error tends toward infinity. Assuming, therefore, that $b<a\delta$, and that at the low-frequency end of the range

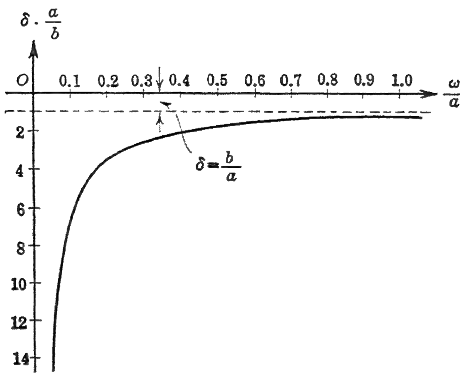


FIG. 30.—The error due to neglecting the line parameter G in the angle of the characteristic impedance function and its dependence upon frequency.

the anti-tangent in (269) is with sufficient accuracy given by its argument, we find that G is negligible in ζ to within a maximum error δ for the range

$$\frac{a\pi}{4}\left\{1 - \sqrt{1 - \frac{16b}{a\pi^2\delta}}\right\} \leq \omega \leq \infty, \tag{270}$$

or if $b \ll a\delta$ we may use the approximate form

$$\frac{2b}{\pi\delta}\left(1 + \frac{4b}{\pi^2 a\delta}\right) \leq \omega \leq \infty. \tag{270a}$$

For the open line data (230) this gives for $\delta = 0.1$

$$\left.\begin{array}{r} 800 \leq \omega \leq \infty \\ 127 \leq f \leq \infty \end{array}\right\}, \tag{271}$$

while for the cable data (233) the range for negligible G is

$$\left.\begin{array}{r} 153 \leq \omega \leq \infty \\ 24 \leq f \leq \infty \end{array}\right\}. \tag{272}$$

Thus under circumstances where L and G are negligible the magnitude of the characteristic impedance may be represented by the simple function (251a) without the $(1 + \delta)$ factor while the angle may be assumed constant and equal to $-\frac{\pi}{4}$ for positive frequencies. For such cases the analytic treatment may, therefore, be considerably simplified.

When circumstances permit only that G be neglected, the analysis may be somewhat simplified by using (264) without the $(1 + \delta)$ factor for $|Z_0|$, and (268) without this factor for the angle ζ. Whenever such approximations are considered, however, it is always necessary first to determine whether, for the essential frequency range involved, the maximum error remains below a value consistent with the degree of accuracy required.

PROBLEMS TO CHAPTER III

3–1. For the open-line data: $R = 10.4$; $L = 0.00367$; $C = 0.00835 \cdot 10^{-6}$; $G = 0.8 \cdot 10^{-6}$, determine and plot the departure functions $f_1(x)$ and $f_2(x)$ as well as the attenuation and phase velocity over a frequency range of 0 to 5000 cycles per second.

3–2. Repeat Problem 3–1 for the cable data: $R = 85.2$; $L = 0.001$; $C = 0.066 \cdot 10^{-6}$; $G = 1.585 \cdot 10^{-6}$.

3–3. For the open-line data of Problem 3–1 determine and plot the magnitude and angle and also the real and imaginary parts of the characteristic impedance function over the range 0–5000 cycles per second.

3–4. Repeat Problem 3–3 for the cable data of Problem 3–2.

3–5. Assuming that in the line of Problem 3–1 the conductance parameter is approximately constant and equal to $0.8 \cdot 10^{-6}$ up to 1000 cycles per second and then increases linearly with frequency at the rate of 10^{-6} mho per 1000 cycles per second, how does this affect the results in Problems 3–1 and 3–3?

3–6. A frequency group of the type shown in Fig. 21 consists of a total of 15 components, the frequency intervals corresponding to $\delta\omega = 500$. Determine and plot the corresponding wave envelope. For a mean frequency corresponding to $\omega_m = 5000$, determine the phase and envelope velocities for the open line of Problem 3–1 and the cable of Problem 3–2.

3–7. The cable whose data are specified in Problem 3–2 is to be loaded so that the maximum departure from constancy on the part of the attenuation function or departure from linearity on the part of the phase function does not exceed 8 per cent over a frequency range of 100 to 3500 cycles per second. Determine approximately the necessary R/L ratio which will meet these specifications. Assuming that the R/L ratio of the loading coils is constant and equal to 80, and that the spacing is sufficiently close so that the loading may be considered approximately uniform, determine the resulting average attenuation and compare with the maximum and minimum attenuation values at the extremes of the frequency range for the unloaded circuit.

3–8. Determine the corresponding approximate maximum departures from ideal behavior in the magnitude and angle of the characteristic impedance function of the loaded circuit of Problem 3–7, assuming the loading to be uniformly distributed; determine and plot the magnitude and angle of Z_0 and compare with the results of Problem 3–4.

3–9. The cable circuit whose data are specified in Problem 3–2 is loaded by inserting coils having a resistance of 10.6 ohms and an inductance of 0.175 henry at intervals of 1.66 miles. Assuming this amount of loading to be uniformly distributed, find the resulting maximum departures from ideal behavior on the part of the attenuation, phase, and characteristic impedance functions over the frequency ranges: $100 < f < 4000$ and $200 < f < 3000$, and compare with the de-

partures found for the corresponding non-loaded facility from the results of Problems 3–1 to 3–4. Also determine and compare the average attenuation and delay per mile of the loaded and unloaded circuits.

3–10. An open telephone line has the following parameter values: $R = 4.14$; $L = 0.00337$; $C = 0.00914 \cdot 10^{-6}$; $G = 0.8 \cdot 10^{-6}$.

(*a*) Determine the range of frequency within which both L and G may be neglected in the propagation function without exceeding an approximate error of 10 per cent.

(*b*) Determine the range of frequencies within which G alone may be neglected in the propagation function to within this same maximum error.

3–11. Consider the statement of Problem 3–10 with reference to the magnitude and angle of the characteristic impedance function.

3–12. Repeat Problem 3–10 for the cable data of Problem 3–2.

3–13. Repeat Problem 3–11 for the cable data of Problem 3–2.

3–14. Reconsider Problem 3–10, assuming the conductance parameter to vary as described in Problem 3–5, and use a cut-and-try procedure to determine how the results are affected.

CHAPTER IV

CHARACTERISTICS OF FOUR-TERMINAL NETWORKS

1. Significance of the problem. In the practical consideration of transmission-line problems we are seldom interested in the distribution of current or voltage along the line, i.e., in what goes on within the structure itself. Primarily the transmission problem concerns itself with the output relative to the input. This was touched upon in sections 15 and 17 of Chapter II where we developed the relations between input and output voltages and currents and discussed some of their ratios. Viewed from this angle, the transmission line or cable is essentially a network having two input and two output terminals, the character of the intervening structure being immaterial except in so far as it determines the behavior of the network as a whole. From the standpoint of the analysis of its external performance we could think of the transmission line as a box with four terminals, two of which are designated as input and two as output terminals. The box itself can be thought of as containing a linear network of some kind whose characteristics are such as to give rise to the external behavior already discussed. This way of viewing the transmission-line problem brings the long line into the same category with a large number of lumped-constant networks which play various important rôles in the complete transmission system as it exists in practice. Such a complete system usually consists not only of the line or cable itself but in addition contains numerous other links whose purpose it is to correct defects in the line performance as well as supply selective or other characteristics to the system which may be required by the class of service which is expected of the facility as a whole.

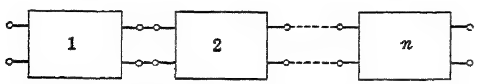


FIG. 31.—Schematic representation of a communication channel consisting of several four-terminal networks in cascade.

A single link of a complete transmission system, which may be thought of as a box with two pairs of terminals, is called a **four-terminal network**, or perhaps more accurately a **two-terminal pair**, since of the four terminals two are always designated as input and two as output terminals. A complete communication channel is made up of a number of such four-terminal networks usually connected in chain or cascade.

Fig. 31 illustrates such a channel in a schematic way. In more complicated systems we may encounter several other types of connection besides the cascade connection of Fig. 31. One or more of the boxes or links may be transmission lines or cables, i.e., distributed-constant networks, while the rest will be lumped-constant networks of various kinds. The performance of the distributed-constant four-terminal networks having been reduced to the same basis as that of the lumped-constant networks, the system as a whole becomes homogeneous and thus we are able to develop a consistent method of analysis for the determination of the overall performance.

In this chapter we wish to focus our attention entirely upon the external behavior of a four-terminal network in order to formulate general concepts as to how this external behavior may be uniquely characterized, how this characterization may be analytically expressed, and how it may be related to the contents of the box. In this connection the effect of various types of symmetry will be treated. Furthermore various types of interconnection of a number of four-terminal networks will be considered. If, in such an interconnection, two pairs of terminals are again designated as input and output, the combination may obviously be looked upon as a single four-terminal network. For example, the cascade arrangement in Fig. 31 may be considered as a single four-terminal network between the input of box 1 and the output of box n. The question arises as to how we are to obtain the resultant performance characteristics of such a system consisting of an interconnection of boxes when their individual properties are known.

Similarly a given four-terminal network may perhaps be more effectively treated if first resolved into simpler component parts each of which is again a four-terminal network. Thus not only the characterization of a given four-terminal network, but its resolution into component parts, as well as its combination with other networks, must be studied before we shall be able to consider a complete transmission system adequately. Incidentally, the results of this analysis will also enable us to treat the situation encountered when line loading is carried out by inserting inductance lumps at intervals, since such a procedure obviously breaks the line up into a cascade of four-terminal units. Problems regarding the design of artificial lines, line balances, filters, and corrective networks of various kinds all require a knowledge of the behavior of four-terminal networks for their attack. Such a study, therefore, seems appropriate at this time.

2. Possible relationships between voltages and currents. Fig. 32 is a sketch of a four-terminal network, indicating by means of arrows the conventions which we shall adopt relative to the reference directions of

currents and voltages at its two ends. The network itself is arbitrary except that it is linear and contains no sources of emf. The convention as to reference directions is a symmetrical one. This is adopted because it does not specify a particular pair of terminals as representing either the input or output end of the network, which fact makes it possible to generalize our analysis more easily. The two pairs of terminals are designated as 1-1′ and 2-2′; these we shall also refer to more briefly as *end 1* and *end 2*.

The behavior of the network with regard to its two pairs of terminals is, therefore, expressible in terms of the four quantities I_1, E_1, I_2, E_2. A linear network having been assumed, it follows that these four quantities must be linearly related. It is also clear from our knowledge of the theory of a lumped-constant network that these relationships may be set up for any two quantities in terms of the other two. There are six such relationships in all, or more specifically, three pairs of mutually inverse relations. Namely, we may express

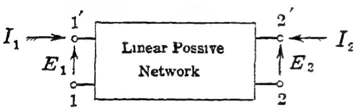


FIG. 32.—Schematic of a four-terminal network showing reference directions for voltage and current.

(a) $\begin{cases} I_1 \text{ and } I_2 \text{ in terms of } E_1 \text{ and } E_2 \\ E_1 \text{ and } E_2 \text{ in terms of } I_1 \text{ and } I_2 \end{cases}$

or

(b) $\begin{cases} I_1 \text{ and } E_2 \text{ in terms of } E_1 \text{ and } I_2 \\ E_1 \text{ and } I_2 \text{ in terms of } I_1 \text{ and } E_2 \end{cases}$

or

(c) $\begin{cases} E_1 \text{ and } I_1 \text{ in terms of } E_2 \text{ and } I_2 \\ E_2 \text{ and } I_2 \text{ in terms of } E_1 \text{ and } I_1. \end{cases}$

All these modes of expression have their particular usefulness in the analysis of four-terminal network performance. We proceed now to derive these and to show their interrelations.

3. Derivation of fundamental relations. If we think of the voltages E_1 and E_2 as forming the closing loops to those meshes of the network which would be left open at the terminals 1-1′ and 2-2′ respectively, and designate these as meshes 1 and 2, then according to the method outlined in Chapter IV of Volume I, and taking due account of the reference directions assumed here, we may write

$$\left.\begin{aligned} I_1 &= \frac{B_{11}}{D} E_1 - \frac{B_{12}}{D} E_2 \\ I_2 &= -\frac{B_{21}}{D} E_1 + \frac{B_{22}}{D} E_2 \end{aligned}\right\}, \tag{273}$$

where D is the network determinant and $B_{ik} = B_{ki}$ are its first minors. Here it is convenient to introduce the notation

$$\left.\begin{aligned} y_{11} &= \frac{B_{11}}{D} \\ y_{12} = y_{21} &= -\frac{B_{12}}{D} \\ y_{22} &= \frac{B_{22}}{D} \end{aligned}\right\}, \tag{274}$$

and then write instead of (273)

$$\left.\begin{aligned} I_1 &= y_{11}E_1 + y_{12}E_2 \\ I_2 &= y_{21}E_1 + y_{22}E_2 \end{aligned}\right\}. \tag{275}$$

If we invert these equations we obtain relations of the form

$$\left.\begin{aligned} E_1 &= z_{11}I_1 + z_{12}I_2 \\ E_2 &= z_{21}I_1 + z_{22}I_2 \end{aligned}\right\}. \tag{276}$$

The z's may be found in terms of the y's by simply solving the system (275) and comparing coefficients with those in (276). Writing for the determinant of (275)

$$|y| = \begin{vmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{vmatrix} = y_{11}y_{22} - y_{12}^2, \tag{277}$$

we find

$$\left.\begin{aligned} z_{11} &= \frac{y_{22}}{|y|} \\ z_{12} = z_{21} &= -\frac{y_{12}}{|y|} \\ z_{22} &= \frac{y_{11}}{|y|} \end{aligned}\right\}. \tag{278}$$

The y's and z's may be interpreted as admittances and impedances of the network. Namely, if we suppose that the terminals 2-2′ are shorted, then obviously E_2 becomes zero, and the equations (275) give

$$\left.\begin{aligned} y_{11} &= \frac{I_1}{E_1} \\ y_{21} &= \frac{I_2}{E_1} \end{aligned}\right\}, \tag{279}$$

while if we imagine the end 1 shorted, then $E_1 = 0$, and we have

$$\left.\begin{aligned} y_{12} &= \frac{I_1}{E_2} \\ y_{22} &= \frac{I_2}{E_2} \end{aligned}\right\}. \tag{280}$$

The admittances y_{11} and y_{22} are evidently those looking into the ends 1 or 2, respectively, with the opposite ends short-circuited. The admittance $y_{21} = y_{12}$ is likewise a short-circuit admittance, but involves the ratio of current at the far end to voltage at the driven end. This we call a **transfer admittance**; the other two are spoken of as **driving-point admittances.** The fact that with alternate ends shorted or driving we have

$$\frac{I_2}{E_1} = \frac{I_1}{E_2}$$

is simply an expression of the reciprocity theorem for this network. The system of y's is referred to as the **short-circuit driving-point and transfer admittances.** These three functions uniquely characterize the behavior of the four-terminal network with respect to its two pairs of terminals.

Similarly, if for the system (276) we regard the end 2 as open, then $I_2 = 0$, and we have

$$\left.\begin{aligned} z_{11} &= \frac{E_1}{I_1} \\ z_{21} &= \frac{E_2}{I_1} \end{aligned}\right\}, \tag{281}$$

while if end 1 is considered open, $I_1 = 0$, and (276) give

$$\left.\begin{aligned} z_{12} &= \frac{E_1}{I_2} \\ z_{22} &= \frac{E_2}{I_2} \end{aligned}\right\}. \tag{282}$$

Here z_{11} and z_{22} are driving-point impedances, and $z_{12} = z_{21}$ is a transfer impedance. The system of z's is, therefore, called the **open-circuit driving-point and transfer impedances.** These are three alternative functions whereby the external behavior of the network may be specified.

The y's or z's may thus be obtained experimentally for any given network. Due attention must, of course, be paid to the convention as to reference directions set down here.

We have still to derive the pairs of relations (b) and (c) referred to in the previous section. These are easily obtained by further algebraic manipulation of any of the systems (273), (275), or (276). By comparison of the coefficients obtained for the various systems we get their interrelationships. We shall not consume space here with the details of this process but merely give the results. Namely, for the pair (b)

we have

$$\left.\begin{aligned} I_1 &= g_{11}E_1 + g_{12}I_2 \\ E_2 &= g_{21}E_1 + g_{22}I_2 \end{aligned}\right\}, \tag{283}$$

and

$$\left.\begin{aligned} E_1 &= h_{11}I_1 + h_{12}E_2 \\ I_2 &= h_{21}I_1 + h_{22}E_2 \end{aligned}\right\}; \tag{284}$$

for the pair (*c*) we find

$$\left.\begin{aligned} E_1 &= \mathcal{A}E_2 - \mathcal{B}I_2 \\ I_1 &= \mathcal{C}E_2 - \mathcal{D}I_2 \end{aligned}\right\}, \tag{285}$$

and

$$\left.\begin{aligned} E_2 &= \mathcal{D}E_1 - \mathcal{B}I_1 \\ I_2 &= \mathcal{C}E_1 - \mathcal{A}I_1 \end{aligned}\right\}. \tag{286}$$

If in addition to (277) we write for the various determinants of these systems

$$|z| = \begin{vmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{vmatrix} = z_{11}z_{22} - z_{12}^2, \tag{287}$$

$$|g| = \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = g_{11}g_{22} - g_{21}g_{12}, \tag{288}$$

$$|h| = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} = h_{11}h_{22} - h_{21}h_{12}, \tag{289}$$

and note that[1]

$$B_{1122} = \frac{1}{D} \cdot \begin{vmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{vmatrix} = \frac{B_{11}B_{22} - B_{12}^2}{D}, \tag{290}$$

then the interrelations between the various sets of coefficients are expressed by

$$\left.\begin{aligned} y_{11} &= \frac{B_{11}}{D} = \frac{z_{22}}{|z|} = \frac{|g|}{g_{22}} = \frac{1}{h_{11}} = \frac{\mathcal{D}}{\mathcal{B}} \\ y_{12} = y_{21} &= -\frac{B_{12}}{D} = -\frac{z_{12}}{|z|} = \frac{g_{12}}{g_{22}} = -\frac{h_{12}}{h_{11}} = -\frac{1}{\mathcal{B}} \\ y_{22} &= \frac{B_{22}}{D} = \frac{z_{11}}{|z|} = \frac{1}{g_{22}} = \frac{|h|}{h_{11}} = \frac{\mathcal{A}}{\mathcal{B}} \end{aligned}\right\}, \tag{291}$$

[1] Here B_{1122} is the minor formed by canceling the first two rows and columns in D. For an independent proof of (290) see M. Bôcher, "Introduction to Higher Algebra," Macmillan & Co., pp. 31–33.

$$\left.\begin{aligned}
z_{11} &= \frac{B_{22}}{B_{1122}} = \frac{y_{22}}{|y|} = \frac{1}{g_{11}} = \frac{|h|}{h_{22}} = \frac{\mathcal{A}}{\mathcal{C}} \\
z_{12} = z_{21} &= \frac{B_{12}}{B_{1122}} = -\frac{y_{12}}{|y|} = -\frac{g_{12}}{g_{11}} = \frac{h_{12}}{h_{22}} = \frac{1}{\mathcal{C}} \\
z_{22} &= \frac{B_{11}}{B_{1122}} = \frac{y_{11}}{|y|} = \frac{|g|}{g_{11}} = \frac{1}{h_{22}} = \frac{\mathcal{D}}{\mathcal{C}}
\end{aligned}\right\}, \tag{292}$$

$$\left.\begin{aligned}
g_{11} &= \frac{B_{1122}}{B_{22}} = \frac{|y|}{y_{22}} = \frac{1}{z_{11}} = \frac{h_{22}}{|h|} = \frac{\mathcal{C}}{\mathcal{A}} \\
g_{12} = -g_{21} &= -\frac{B_{12}}{B_{22}} = \frac{y_{12}}{y_{22}} = -\frac{z_{12}}{z_{11}} = -\frac{h_{12}}{|h|} = -\frac{1}{\mathcal{A}} \\
g_{22} &= \frac{D}{B_{22}} = \frac{1}{y_{22}} = \frac{|z|}{z_{11}} = \frac{h_{11}}{|h|} = \frac{\mathcal{B}}{\mathcal{A}}
\end{aligned}\right\}, \tag{293}$$

$$\left.\begin{aligned}
h_{11} &= \frac{D}{B_{11}} = \frac{1}{y_{11}} = \frac{|z|}{z_{22}} = \frac{g_{22}}{|g|} = \frac{\mathcal{B}}{\mathcal{D}} \\
h_{12} = -h_{21} &= \frac{B_{12}}{B_{11}} = -\frac{y_{12}}{y_{11}} = \frac{z_{12}}{z_{22}} = -\frac{g_{12}}{|g|} = \frac{1}{\mathcal{D}} \\
h_{22} &= \frac{B_{1122}}{B_{11}} = \frac{|y|}{y_{11}} = \frac{1}{z_{22}} = \frac{g_{11}}{|g|} = \frac{\mathcal{C}}{\mathcal{D}}
\end{aligned}\right\}, \tag{294}$$

$$\left.\begin{aligned}
\mathcal{A} &= \frac{B_{22}}{B_{12}} = -\frac{y_{22}}{y_{12}} = \frac{z_{11}}{z_{12}} = -\frac{1}{g_{12}} = \frac{|h|}{h_{12}} \\
\mathcal{B} &= \frac{D}{B_{12}} = -\frac{1}{y_{12}} = \frac{|z|}{z_{12}} = -\frac{g_{22}}{g_{12}} = \frac{h_{11}}{h_{12}} \\
\mathcal{C} &= \frac{B_{1122}}{B_{12}} = -\frac{|y|}{y_{12}} = \frac{1}{z_{12}} = -\frac{g_{11}}{g_{12}} = \frac{h_{22}}{h_{12}} \\
\mathcal{D} &= \frac{B_{11}}{B_{12}} = -\frac{y_{11}}{y_{12}} = \frac{z_{22}}{z_{12}} = -\frac{|g|}{g_{12}} = \frac{1}{h_{12}}
\end{aligned}\right\}. \tag{295}$$

In addition to these it may be useful to recognize that

$$\left.\begin{aligned}
|y| = |z|^{-1} &= \frac{B_{1122}}{D} = \frac{g_{11}}{g_{22}} = \frac{h_{22}}{h_{11}} = \frac{\mathcal{C}}{\mathcal{B}} \\
|g| = |h|^{-1} &= \frac{B_{11}}{B_{22}} = \frac{y_{11}}{y_{22}} = \frac{z_{22}}{z_{11}} = \frac{\mathcal{D}}{\mathcal{A}}
\end{aligned}\right\}. \tag{296}$$

The pair of relations (285) and (286) have for many years played an important rôle in the field of power transmission over long lines. There the notation of $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ for the coefficients was first introduced, and they are spoken of in power work as general circuit constants. In communication work, where the behavior of a system versus frequency is the important factor, we prefer to speak of them as the **general circuit parameters.**

We retain the notation $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ for the coefficients in (285) and

(286) because of the fact that this has become common usage in the literature. Due to the relation

$$\begin{vmatrix} \mathfrak{A} & \mathfrak{B} \\ \mathfrak{C} & \mathfrak{D} \end{vmatrix} = \mathfrak{A}\mathfrak{D} - \mathfrak{B}\mathfrak{C} = 1, \tag{297}$$

which follows from (295), the inverse system (286) has for its coefficients also simply the letters $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, $\mathfrak{D}$. Hence retaining this form of notation introduces no particular hardship for us, although a new notation consistent with that used in the other systems of equations would have given more uniformity to the series of systems as a whole.

Regarding the other systems, (275), (276), (283), and (284), there is as yet no tradition either as to form or notation. We have, therefore, chosen these as we thought best.

The student should note that, no matter how the external behavior of the four-terminal network is expressed, *three independent functions of frequency always suffice to do this.*

4. Symmetrical networks. A four-terminal network is said to be symmetrical with respect to its ends 1 and 2 when an interchange of these pairs of terminals has no effect upon the external behavior. Since the equations (285) and (286) relate the voltage and current of one end to those at the other, they afford a good opportunity for investigating the necessary conditions for the existence of such symmetry. Here we see that these equations will become interchanged for an interchange of the subscripts 1 and 2 if

$$\mathfrak{A} = \mathfrak{D}. \tag{298}$$

This, therefore, becomes the symmetry condition. Referring to (295) we see that this condition can be put into the following alternative forms

$$\left.\begin{aligned} B_{11} &= B_{22} \\ y_{11} &= y_{22} \\ z_{11} &= z_{22} \\ |g| &= |h| = 1 \end{aligned}\right\}. \tag{298a}$$

The first of these is interesting. It shows that the internal structure of the network need not be symmetrical in order for the external behavior to be symmetrical. This becomes clear if we recognize that $B_{11} = B_{22}$ does not require symmetry with respect to all of the internal meshes of the network. It is sufficient that the first two rows and the first two columns of the network determinant be equal, although this may not even be necessary. The remainder of this determinant may have any form.

Note that, when the symmetry condition is satisfied, only two independent functions remain in any of the above sets (291) to (295). Hence we see that *a symmetrical four-terminal network is uniquely characterized by two functions.* The long line is evidently a symmetrical system on account of its uniformity. It will be recalled that in section 16 of Chapter II we pointed out that the line behavior per se is uniquely characterized by two functions which may be either Z_0 and α or Z_{oc} and Z_{sc}, the latter being the same as our present functions $z_{11} = z_{22}$ and $1/y_{11} = 1/y_{22}$, respectively.

5. The use of matrix algebra in the study of four-terminal network behavior. The manipulations involved in the application of the above six transformations become simplified if we make use of matrix algebra. A **matrix** has an appearance similar to that of a determinant. Like the latter, it is a rectangular arrangement of coefficients in the form of rows and columns. Unlike a determinant, a matrix need not have the same number of rows as columns, nor does it possess a value which is arrived at by some process of expansion or other. It is simply a rectangular array of elements, and nothing more.

Matrix algebra was developed for the purpose of simplifying the manipulation of several systems of linear simultaneous equations or linear transformations, just as determinant algebra was developed in order to simplify or organize the process of solution of such systems. The various rules for the addition or multiplication, etc., of matrices were not arbitrarily set down, but determined out of the resulting forms obtained by performing these operations upon the systems of equations themselves. Thus matrix algebra is nothing more than a collection of rules for manipulating the coefficients of simultaneous equations when various sets of such equations are to be added or substituted into one another, and so on.

Consider, for example, the system of equations

$$\left.\begin{aligned} a_{11}\xi_1 + a_{12}\xi_2 &= \eta_1 \\ a_{21}\xi_1 + a_{22}\xi_2 &= \eta_2 \end{aligned}\right\}. \tag{299}$$

The following arrangement of coefficients

$$\begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix} \tag{300}$$

is called the matrix of this system, or, since (299) is sometimes spoken of as a linear transformation (transforming the ξ's into η's), (300) is also referred to as the transformation matrix, or the matrix of the transformation. The term "substitution" is also used in place of "transformation."

Now suppose that the η's of (299) are transformed into ζ's by the system.

$$\left.\begin{aligned} b_{11}\eta_1 + b_{12}\eta_2 &= \zeta_1 \\ b_{21}\eta_1 + b_{22}\eta_2 &= \zeta_2 \end{aligned}\right\}, \tag{301}$$

and it is desired to find how the ξ's in (299) are related to the ζ's of (301). To determine this we must substitute the η's of (299) into (301). and then arrange the terms in the order in which they appear in (299), If we do this we will find that the desired relations are

$$\left.\begin{aligned} c_{11}\xi_1 + c_{12}\xi_2 &= \zeta_1 \\ c_{21}\xi_1 + c_{22}\xi_2 &= \zeta_2 \end{aligned}\right\}, \tag{302}$$

where

$$\left.\begin{aligned} c_{11} &= a_{11}b_{11} + a_{12}b_{21} \\ c_{12} &= a_{11}b_{12} + a_{12}b_{22} \\ c_{21} &= a_{21}b_{11} + a_{22}b_{21} \\ c_{22} &= a_{21}b_{12} + a_{22}b_{22} \end{aligned}\right\}. \tag{303}$$

The student should work this out on paper and check the relations (303) in order to appreciate the steps involved. He will then also appreciate that the mechanical work involved will become quite heavy in systems with many equations and many unknowns. This labor can now be materially shortened if we recognize certain characteristics in the form of the coefficients (303).

Consider in addition to the matrix (300), the corresponding one for (301) which is

$$\begin{Vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{Vmatrix}. \tag{304}$$

Now observe that c_{11} in (303) is simply the product of the *first row* in (300) with the *first column* in (304), formed in the same way in which rows or columns are multiplied by each other when writing the product of two determinants, i.e., the first element in a row is multiplied by the first element in a column, the second element in the row by the second element in the column, and so on, and the results added. The coefficient c_{12} is then the product of the *first row* of (300) and the *second column* of (304). Finally c_{21} and c_{22} are products of the *second row* in (300) and the *first* and *second columns* respectively in (304). The first index in the c's refers to the row in (300), and the second index to the column in (304). This is precisely the way in which products of

determinants are formed, as the reader will probably recall. Hence we say that the matrix

$$\begin{Vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{Vmatrix} \tag{305}$$

is the product of (300) and (304). This is indicated as follows

$$\|c\| = \|a\| \times \|b\|. \tag{306}$$

Note, however, that whereas in forming the product of two determinants it makes no difference in the value of the resulting determinant whether we multiply rows of the first by columns of the second, or rows by rows, or columns by rows, or columns by columns, there is only *one* way of multiplying matrices and that is the way in which the result (303) prescribes, namely by multiplying the *rows of the first by the columns of the second.* This fact also shows that the multiplication of matrices does not follow the commutative law, i.e.,

$$\|a\| \times \|b\| \neq \|b\| \times \|a\|,$$

as the reader may easily prove for himself.

Thus the substitution of (299) into (301) gives rise to (302), whose matrix is the product of those of (299) and (301). This together with the rule for multiplying matrices should simplify for us any such processes of substitution which we shall have to make in the future.

On the basis of the multiplication rule just developed, we can write the system (299) in matrix form, thus

$$\begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix} \times \begin{Vmatrix} \xi_1 \\ \xi_2 \end{Vmatrix} = \begin{Vmatrix} \eta_1 \\ \eta_2 \end{Vmatrix}. \tag{299a}$$

The second matrix on the left has only one column. Hence the product matrix has also only one column; and the equality of this with the right-hand matrix means that corresponding elements in these two are equal, for *two matrices are equal only when all corresponding elements are equal.* Thus the matrix equation (299*a*) is equivalent to (299). Similarly (301) would be in matrix form

$$\begin{Vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{Vmatrix} \times \begin{Vmatrix} \eta_1 \\ \eta_2 \end{Vmatrix} = \begin{Vmatrix} \zeta_1 \\ \zeta_2 \end{Vmatrix}. \tag{301a}$$

The substitution of (299) in (301) is in matrix form obtained from (299*a*) and (301*a*), namely

$$\begin{Vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{Vmatrix} \times \begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix} \times \begin{Vmatrix} \xi_1 \\ \xi_2 \end{Vmatrix} = \begin{Vmatrix} \zeta_1 \\ \zeta_2 \end{Vmatrix}, \tag{307}$$

and by (306) we then have

$$\begin{Vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{Vmatrix} \times \begin{Vmatrix} \xi_1 \\ \xi_2 \end{Vmatrix} = \begin{Vmatrix} \zeta_1 \\ \zeta_2 \end{Vmatrix}. \tag{302a}$$

It might be well for the reader to acquaint himself with further details concerning matrix algebra.[1] We shall give here only those few operations with matrices which are absolutely essential to our purpose.

A few more facts about matrix manipulation are necessary before we can proceed with our network problem. The process of solving simultaneous equations determines the significance of the reciprocal or *inverse* of a matrix. Namely, if we consider (299*a*) and ask for the solution, i.e., the ξ's in terms of the η's, then it seems reasonable that this should be given by

$$\begin{Vmatrix} \xi_1 \\ \xi_2 \end{Vmatrix} = \begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix}^{-1} \times \begin{Vmatrix} \eta_1 \\ \eta_2 \end{Vmatrix}. \tag{308}$$

The matrix $\|a\|^{-1}$ is called the *inverse* of $\|a\|$. The formation of the inverse matrix from the original must be carried out in such a way that (308) will be valid. The determinant method of solving simultaneous equations gives the answer. Namely, the inverse matrix is formed by first replacing each element by its minor in the corresponding determinant, dividing each minor by this determinant, and then interchanging rows and columns. Denoting the corresponding determinant of $\|a\|$ by

$$|a| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \tag{309}$$

we have by applying this rule in the above case

$$\begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix}^{-1} = \begin{Vmatrix} \dfrac{a_{22}}{|a|} & -\dfrac{a_{12}}{|a|} \\ -\dfrac{a_{21}}{|a|} & \dfrac{a_{11}}{|a|} \end{Vmatrix}. \tag{310}$$

Since the determinant here has only two rows and columns, the minors are single elements.

A matrix is multiplied by a factor when each of its elements is multiplied by this factor, i.e.,

$$k \cdot \begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix} = \begin{Vmatrix} ka_{11} & ka_{12} \\ ka_{21} & ka_{22} \end{Vmatrix}. \tag{311}$$

[1] See for example M. Bôcher, "Introduction to Higher Algebra," Macmillan Co., 1927

Note the difference between this and the multiplication of a determinant by a factor. In that case only one row or column contains this factor. From (311) we see that a matrix changes sign only when all its elements change sign; and it vanishes only when all its elements are zero.

Finally, *matrices are added by adding corresponding elements*, thus

$$\begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix} + \begin{Vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{Vmatrix} = \begin{Vmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{Vmatrix}. \tag{312}$$

This follows from the addition of systems of equations involving the same set of unknowns.

With this brief excursion into matrix algebra we return to our four-terminal network transformations.[1] In matrix form these are

$$\begin{Vmatrix} I_1 \\ I_2 \end{Vmatrix} = \begin{Vmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{Vmatrix} \times \begin{Vmatrix} E_1 \\ E_2 \end{Vmatrix}, \tag{275a}$$

$$\begin{Vmatrix} E_1 \\ E_2 \end{Vmatrix} = \begin{Vmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{Vmatrix} \times \begin{Vmatrix} I_1 \\ I_2 \end{Vmatrix}, \tag{276a}$$

$$\begin{Vmatrix} I_1 \\ E_2 \end{Vmatrix} = \begin{Vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{Vmatrix} \times \begin{Vmatrix} E_1 \\ I_2 \end{Vmatrix}, \tag{283a}$$

$$\begin{Vmatrix} E_1 \\ I_2 \end{Vmatrix} = \begin{Vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{Vmatrix} \times \begin{Vmatrix} I_1 \\ E_2 \end{Vmatrix}, \tag{284a}$$

$$\begin{Vmatrix} E_1 \\ I_1 \end{Vmatrix} = \begin{Vmatrix} \mathfrak{A} & \mathfrak{B} \\ \mathfrak{C} & \mathfrak{D} \end{Vmatrix} \times \begin{Vmatrix} E_2 \\ -I_2 \end{Vmatrix}, \tag{285a}$$

$$\begin{Vmatrix} E_2 \\ I_2 \end{Vmatrix} = \begin{Vmatrix} \mathfrak{D} & \mathfrak{B} \\ \mathfrak{C} & \mathfrak{A} \end{Vmatrix} \times \begin{Vmatrix} E_1 \\ -I_1 \end{Vmatrix}. \tag{286a}$$

Taken in pairs, these transformations are each others' inverses, and their respective matrices are inverse. The utility of these matrix forms lies primarily in finding the resulting substitutions for several four-terminal networks which are interconnected in various ways. By transforming from one to another, we may also obtain equivalent networks.

[1] The application of matrix algebra to the general treatment of the four-terminal network was first given by F. Strecker and R. Feldtkeller, "Grundlagen der Theorie des allgemeinen Vierpols," E.N.T. 6, pp. 93–112, 1929. See also H. G. Baerwald, "Die Eigenschaften Symmetrischer 4n-Pole . . .", Sitzb. d. Preuss. Akad. d. Wiss. Phys.-Math. Kl. 33, pp. 784–829, 1931.

Considering two four-terminal networks, these may be interconnected in the following five fundamental ways, namely:

(a) cascade
(b) parallel
(c) series
(d) series-parallel
(e) parallel-series.

Fig. 33 illustrates this. In the following we shall treat these fundamental cases only. When more component networks are involved, an extension of the same methods may be applied.

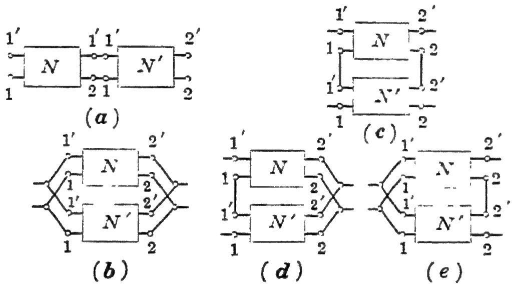


FIG. 33.—Possible interconnections of a pair of dissimilar four-terminal networks.

(a) *Cascade connection.* For the first network in the chain let us assume that the input in terms of the output is given by (285*a*). For the second we will denote the voltages and currents as well as the coefficients of the matrix by the same letters primed, thus

$$\begin{Vmatrix} E_1' \\ I_1' \end{Vmatrix} = \begin{Vmatrix} \mathfrak{A}' & \mathfrak{B}' \\ \mathfrak{C}' & \mathfrak{D}' \end{Vmatrix} \times \begin{Vmatrix} E_2' \\ -I_2' \end{Vmatrix}. \tag{285b}$$

Since the output of the first network equals the input to the second

$$E_2 = E_1'; \ -I_2 = I_1'.$$

Hence, substituting (285*b*) into (285*a*), we have

$$\begin{Vmatrix} E_1 \\ I_1 \end{Vmatrix} = \begin{Vmatrix} \mathfrak{A} & \mathfrak{B} \\ \mathfrak{C} & \mathfrak{D} \end{Vmatrix} \times \begin{Vmatrix} \mathfrak{A}' & \mathfrak{B}' \\ \mathfrak{C}' & \mathfrak{D}' \end{Vmatrix} \times \begin{Vmatrix} E_2' \\ -I_2' \end{Vmatrix}, \tag{313}$$

which gives the input in terms of the output for the two networks in cascade. The resultant transformation matrix is obviously obtained by forming the product of the two individual matrices.

(*b*) *Parallel connection.* Here the voltages at the ends of the two networks are common. Using (275*a*) for the top network, and

$$\begin{Vmatrix} I_1' \\ I_2' \end{Vmatrix} = \begin{Vmatrix} y_{11}' & y_{12}' \\ y_{21}' & y_{22}' \end{Vmatrix} \times \begin{Vmatrix} E_1' \\ E_2' \end{Vmatrix} \tag{275b}$$

for the bottom one, and noting that

$$E_1 = E_1';\ E_2 = E_2',$$

we get by adding these matrix equations according to (312)

$$\begin{Vmatrix} I_1 + I_1' \\ I_2 + I_2' \end{Vmatrix} = \begin{Vmatrix} y_{11} + y_{11}' & y_{12} + y_{12}' \\ y_{21} + y_{21}' & y_{22} + y_{22}' \end{Vmatrix} \times \begin{Vmatrix} E_1 \\ E_2 \end{Vmatrix} \tag{314}$$

for the resulting parallel connection. The coefficients of the resultant matrix are simply the sums of those for the individual matrices. By applying the relations (292) to (295) the resultant y-system may be converted into any of the others.

(*c*) *Series connection.* Here we use (276*a*) for the top network and

$$\begin{Vmatrix} E_1' \\ E_2' \end{Vmatrix} = \begin{Vmatrix} z_{11}' & z_{12}' \\ z_{21}' & z_{22}' \end{Vmatrix} \times \begin{Vmatrix} I_1' \\ I_2' \end{Vmatrix} \tag{276b}$$

for the bottom one. Then, since

$$I_1 = I_1';\ I_2 = I_2',$$

we have by adding (276*a*) and (276*b*)

$$\begin{Vmatrix} E_1 + E_1' \\ E_2 + E_2' \end{Vmatrix} = \begin{Vmatrix} z_{11} + z_{11}' & z_{12} + z_{12}' \\ z_{21} + z_{21}' & z_{22} + z_{22}' \end{Vmatrix} \times \begin{Vmatrix} I_1 \\ I_2 \end{Vmatrix} \tag{315}$$

for the series connection. Here the resultant transformation matrix is obtained by adding coefficients in the individual z-systems. The resulting z-system may, of course, be converted into any of the others if this should be desired.

(*d*) *Series-parallel connection.* For the treatment of this case we use the transformation (284*a*) because here the input currents and the output voltages are common. Thus if we let (284*a*) be the transformation for one network, and

$$\begin{Vmatrix} E_1' \\ I_2' \end{Vmatrix} = \begin{Vmatrix} h_{11}' & h_{12}' \\ h_{21}' & h_{22}' \end{Vmatrix} \times \begin{Vmatrix} I_1' \\ E_2' \end{Vmatrix} \tag{284b}$$

that for the other, and note that

$$I_1 = I_1';\ E_2 = E_2',$$

we have by adding

$$\begin{Vmatrix} E_1 + E_1' \\ I_2 + I_2' \end{Vmatrix} = \begin{Vmatrix} h_{11} + h_{11}' & h_{12} + h_{12}' \\ h_{21} + h_{21}' & h_{22} + h_{22}' \end{Vmatrix} \times \begin{Vmatrix} I_1 \\ E_2 \end{Vmatrix}, \tag{316}$$

which is the transformation for the series-parallel combination.

(*e*) *Parallel-series connection.* Here (283*a*) together with

$$\begin{Vmatrix} I_1' \\ E_2' \end{Vmatrix} = \begin{Vmatrix} g_{11}' & g_{12}' \\ g_{21}' & g_{22}' \end{Vmatrix} \times \begin{Vmatrix} E_1' \\ I_2' \end{Vmatrix} \tag{283b}$$

are used. Since in this connection

$$E_1 = E_1'; I_2 = I_2',$$

addition of the matrix equations gives

$$\begin{Vmatrix} I_1 + I_1' \\ E_2 + E_2' \end{Vmatrix} = \begin{Vmatrix} g_{11} + g_{11}' & g_{12} + g_{12}' \\ g_{21} + g_{21}' & g_{22} + g_{22}' \end{Vmatrix} \times \begin{Vmatrix} E_1 \\ I_2 \end{Vmatrix} \tag{317}$$

as the desired relation for the combined system.

Various combinations of these fundamental modes of connection may be treated by applying the same general principles. There is one important restriction to the method, however, which should be pointed out at this time. This has to do with the cases *b*, *c*, *d*, and *e*. When one four-terminal network operates by itself, it is evident that the current which enters terminal 1′ (Fig. 32) is identical with that which emerges from terminal 1. Both are I_1. Likewise the current I_2 is that which enters 2′ or emerges from 2. When several networks are inter-connected, however, this condition is not assured unless certain other requirements are satisfied. If the current entering 1′ is not the same as that leaving 1, or that entering 2′ is not the same as that leaving 2, the individual transformations which apply for each network by itself no longer hold when the networks are interconnected, and hence the above methods of determining the combined performance fail![1]

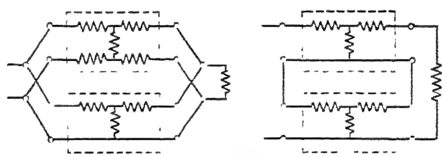


FIG. 34.—Parallel and series inter-connections of four-terminal networks for which the matrix method of determining the composite behavior does not apply.

Fig. 34 illustrates a parallel and a series case for which this condition

[1] H. G. Baerwald, "Der Geltungsbereich der Strecker-Feldtkellerschen Matrizengleichungen von Vierpolsystemen," E.N.T., 9, p. 31, 1932.

exists, i.e., for which the above methods of combination are invalid. The student should convince himself of this by assigning arbitrary finite non-zero resistance values to these networks and calculating the currents entering the terminals of the individual networks for an assumed impressed voltage.

The validity of the above methods of combination may be tested by applying simple rules for the individual cases.[1] These rules are based upon a recognition of the cause of the current unbalance at the pairs of network terminals. Consider the parallel connection of Fig. 33*b*. Suppose the networks are connected in parallel on their left-hand sides, but that they are individually terminated in such impedances that the voltages E_2 and E_2' are equal. The right-hand sides can then be placed in parallel without disturbing the individual behaviors in any way *provided no potential difference exists between the terminals to be joined,* for, if potential differences do exist between these terminals, then currents will circulate between the networks after the connection is made, and thus the combined behavior will not be given by a superposition of the previous individual behaviors.

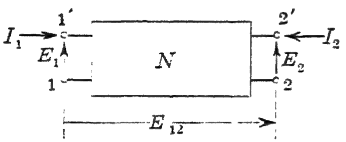


FIG. 35.—Terminal voltage considerations necessary for the determination of the validity conditions for the parallel interconnection when treated by matrix algebra.

Furthermore, this requirement must be met for *all frequencies* and for *all possible load conditions.*

In order to formulate this more definitely we must determine these additional potential differences for the connection in question and see under what circumstances they will be zero. In Fig. 35 we have shown, in addition to the usual voltages and currents, the voltage E_{12} which appears between the terminals 1 and 2. By applying the usual principles of lumped network theory, this voltage may be expressed in terms of the voltages E_1 and E_2. Suppose we write for the network N

$$E_{12} = a_1E_1 + a_2E_2, \tag{318}$$

and for the network to be placed in parallel

$$E_{12}' = a_1'E_1' + a_2'E_2'. \tag{318a}$$

Since, for this connection we must have $E_1 = E_1'$ and $E_2 = E_2'$, E_{12} will simultaneously be equal to E_{12}' if

$$a_1 = a_1'; a_2 = a_2'. \tag{319}$$

If this condition (319) is fulfilled, then E_{12} will equal E_{12}' for all loads

[1] O. Brune, E.N.T., Vol. 9, No. 6, p. 234, 1932.

and for all frequencies simultaneously with the condition $E_1 = E_1'$, $E_2 = E_2'$. Then the voltages between the terminals 1′ and 2′ will also be equal in both networks, and consequently there will be no circulatory currents between them after the parallel connection is made. The relations (319) are, therefore, the necessary and sufficient conditions for the validity of the matrix method of determining the combined behavior for the parallel connection.

These conditions may now be given a physical interpretation whereby they are more easily applied to a specific case. Suppose the networks are connected in parallel on their input sides but individually short-circuited on their output sides, as shown in Fig. 36*a*. Then $E_2 = E_2' = 0$, so that the condition $a_1 = a_1'$ for (318) and (318*a*) becomes $E_{12} = E_{12}'$ which, for the existing connections, is the same as $V = 0$ where V is the voltage appearing between the short-circuited ends. If both networks are reversed, the test for $V = 0$ will correspond to $a_2 = a_2'$. These *two* tests, which for most practical cases may be carried out by inspection, take the place of the conditions (319).

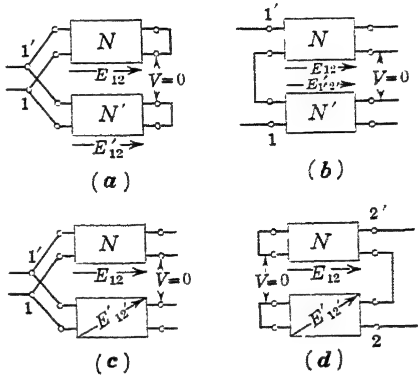


FIG. 36.—Validity tests for matrix methods of determining composite behavior.

For the series connection the validity test is illustrated by Fig. 36*b*. Here we can write

$$\left.\begin{aligned} E_{12} &= b_1 I_1 + b_2 I_2 \\ E_{1'2'} &= b_1' I_1' + b_2' I_2' \end{aligned}\right\}, \tag{320}$$

and since for the series connection it is necessary that

$$I_1 = I_1';\ I_2 = I_2'$$

shall occur simultaneously with

$$E_{12} = E_{1'2'},$$

the necessary and sufficient conditions become

$$b_1 = b_1';\ b_2 = b_2'. \tag{321}$$

The first of these conditions may be checked by having the left-hand sides connected in series while the right-hand sides are open so that $I_2 = I_2' = 0$. Then $V = 0$ corresponds to $b_1 = b_1'$ as shown in Fig. 36*b*. Repeating the test with the networks reversed is a test for $b_2 = b_2'$.

Finally the tests for the parallel-series connection are shown in Fig.

36*c* and *d*. Here the voltages which should be equal for all loads and all frequencies simultaneously with $E_1 = E_1'$ and $I_2 = I_2'$ are E_{12} and $E_{12'}'$. For these we can write

$$\left.\begin{aligned} E_{12} &= c_1E_1 + c_2I_2 \\ E_{12'}' &= c_1'E_1' + c_2'I_2' \end{aligned}\right\}, \tag{322}$$

and thus have for the validity conditions

$$c_1 = c_1';\ c_2 = c_2'. \tag{322a}$$

For the connection of Fig. 36*c* we have $I_2 = I_2' = 0$; and since $V = 0$ corresponds to $E_{12} = E_{12'}'$, we see that this is a test for the first condition (322*a*). The connection of Fig. 36*d*, on the other hand, makes $E_1 = E_1' = 0$ and $I_2 = I_2'$ so that $V = 0$ is the test for the second condition (322*a*). The series-parallel connection is essentially the same as this and, therefore, needs no special comment.

The student may apply these rules or tests to the network combinations of Fig. 34. In these instances it is quite obvious that they are not satisfied. As an alternative to this method of testing by inspection of the network, he should calculate the coefficients a_1, a_2, b_1, and b_2 of equations (318) and (320) for these same networks and then apply the criteria (319) and (321). In this way he will appreciate how much simpler it is to apply the tests by inspection than to make the corresponding validity check analytically. When the networks involved are fairly complicated, it may not be possible to apply the tests by inspection. Then the coefficients in the systems (318), (320), and (322) must be determined. This may be quite laborious but it can always be done by the application of usual network principles.

When a larger number of networks are involved, the rules or tests are carried out in the same way. Ends to be paralleled or placed in series are short-circuited or left open respectively, while the opposite ends are connected in the desired manner, and the voltages determined between all terminals which are to be joined. The networks are then reversed and the tests repeated. All these voltages must vanish at all frequencies in order for the matrix method of combination to be valid.

In network synthesis, i.e., in the design of four-terminal networks which are to meet certain prescribed characteristics, we may find that the desired result can be obtained from the combination of several component networks after the matrix fashion. The problem may then be considered solved provided the network combination satisfies the necessary tests for the particular interconnection involved. When these are not satisfied, the difficulty may be overcome in one of several ways.

Cases like those illustrated in Fig. 34 are very simple to handle. In

the parallel connection, for example, if we modify the upper network by removing the series resistances from its bottom branches and adding these to the series resistances of its top branches, respectively, it will have the same structure as the lower network, and its external behavior will be unchanged. In this modified form it may be placed in parallel with the other network of similar form without violating the conditions for this connection.

For the series combination of Fig. 34 the situation is still simpler. Here the conditions for the connection are met by inverting the lower network so that the result becomes symmetrical about the horizontal center line.

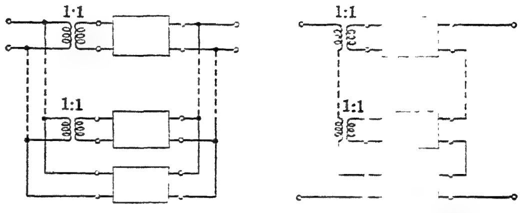


FIG. 37.—Use of ideal transformers in network interconnections for obtaining composite behavior as predicted by matrix methods when validity conditions are not satisfied.

When these or similar measures are insufficient, the situation can always be met by using either input or output transformers of a 1:1 ratio in conjunction with the individual networks. In this way the currents entering and leaving the ends of each network are *forced* to be equal. The transformers must of course be *ideal* so as not to affect the net behavior. This introduces a difficulty from the practical standpoint which can be only approximately met. It should also be pointed out in this connection that transformers need be used only on all but one of the component networks. In two networks, therefore, only one of these needs to be connected through an ideal transformer. Fig. 37 illustrates the general scheme of transformer and network connections for the parallel and the series cases.

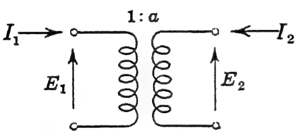


FIG. 38.—Schematic of an ideal transformer considered as a two-terminal pair.

6. Ideal transformers and transformers without loss. Since ideal transformers of various ratios enter into the discussion of four-terminal network analysis quite frequently, it might be well to point out at this time how their characteristics are taken into account.

For the ideal transformer of ratio $1\!:\!a$, illustrated in Fig. 38, the following system of equations may be written

$$\left.\begin{aligned} E_1 &= \frac{1}{a}E_2 - 0I_2 \\ I_1 &= 0E_2 - aI_2 \end{aligned}\right\} \qquad (323)$$

in which the terms with zero coefficients are included in order to make it clear that this is a special form of (285). In matrix form we have

$$\begin{Vmatrix} E_1 \\ \\ I_1 \end{Vmatrix} = \begin{Vmatrix} \frac{1}{a} & 0 \\ \\ 0 & a \end{Vmatrix} \times \begin{Vmatrix} E_2 \\ \\ -I_2 \end{Vmatrix}. \tag{323a}$$

For the $\mathfrak{ABCD}$-system of the ideal transformer, therefore, we get

$$\mathfrak{A}_t = \frac{1}{a}; \ \mathfrak{B}_t = 0; \ \mathfrak{C}_t = 0; \ \mathfrak{D}_t = a. \tag{324}$$

It frequently occurs that a given four-terminal network works out of or into an ideal transformer, and it is desired to determine the combined behavior. This is simply a cascade connection of the transformer and the given network. If the network is characterized by $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, $\mathfrak{D}$, then the $\mathfrak{ABCD}$-matrix for the combination will be

$$\begin{Vmatrix} \frac{1}{a} & 0 \\ \\ 0 & a \end{Vmatrix} \times \begin{Vmatrix} \mathfrak{A} & \mathfrak{B} \\ \\ \mathfrak{C} & \mathfrak{D} \end{Vmatrix} = \begin{Vmatrix} \frac{\mathfrak{A}}{a} & \frac{\mathfrak{B}}{a} \\ \\ a\mathfrak{C} & a\mathfrak{D} \end{Vmatrix} \tag{325}$$

when the transformer *precedes* the network, and

$$\begin{Vmatrix} \mathfrak{A} & \mathfrak{B} \\ \\ \mathfrak{C} & \mathfrak{D} \end{Vmatrix} \times \begin{Vmatrix} \frac{1}{a} & 0 \\ \\ 0 & a \end{Vmatrix} = \begin{Vmatrix} \frac{\mathfrak{A}}{a} & a\mathfrak{B} \\ \\ \frac{\mathfrak{C}}{a} & a\mathfrak{D} \end{Vmatrix} \tag{326}$$

when the transformer *follows* the network.

These resulting $\mathfrak{ABCD}$-systems may be transformed into any of the others by using the relations (291) to (294). For example, if $\|y\|$ and $\|z\|$ are the matrices for the network alone, then when the transformer precedes the network we have

$$\begin{Vmatrix} a^2y_{11} & ay_{12} \\ \\ ay_{21} & y_{22} \end{Vmatrix}, \tag{327}$$

and

$$\begin{Vmatrix} \frac{z_{11}}{a^2} & \frac{z_{12}}{a} \\ \\ \frac{z_{21}}{a} & z_{22} \end{Vmatrix} \tag{328}$$

for the overall matrices in the y- and z-systems respectively.

These might have been determined by inspection from the diagram of Fig. 39. For example, the admittance looking into end 1 of the

network with end 2 shorted is y_{11}. On the left side of the ideal transformer, however, this looks like a^2y_{11}. Looking into end 2 of the network with the left side of the transformer short-circuited, we simply see y_{22} because the ideal transformer constitutes a short circuit when one of its sides is shorted. For the transfer admittance of the combination, suppose end 2 is short-circuited and the ratio of current at that end determined relative to the voltage impressed upon the transformer. Since the transformer steps up the voltage in the ratio of $1:a$, it is clear that this transfer admittance must be ay_{12}, as found by the other method. In the same way the z-matrix (328) can also be determined by inspection.

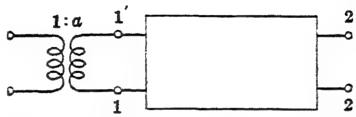


Fig. 39.—Ideal transformer preceding a two-terminal pair.

Since an ideal transformer is one which has neither loss nor leakage and in addition has infinite inductances, it is rather difficult to approximate physically. Although in actual transformers the losses may be kept within reason, and the inductive reactances made large compared with those with which the transformer is associated, it is frequently difficult to keep the leakage low, especially with high transformation ratios or in air-core transformers. In such cases it is sometimes possible to absorb the leakage inductance of the transformer, either wholly or partly, in an adjacent inductance belonging to the network with which the transformer is associated.

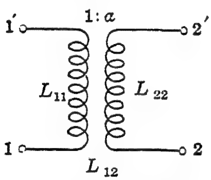


Fig. 40.—Non-dissipative transformer viewed as a two-terminal pair.

In order to illustrate how this may be done we shall consider a transformer having no loss but otherwise physically realizable except for extraneous capacitances. The winding resistances, which are responsible for the loss, are omitted in this discussion merely to simplify the treatment. They may easily be introduced subsequently without altering anything in the following argument. Fig. 40 illustrates the circuit and also gives the notation for the self- and mutual inductances.

By inspection we can write for this network

$$z_{11} = jL_{11}\omega;\ z_{12} = jL_{12}\omega;\ z_{22} = jL_{22}\omega,$$

where ω represents the angular frequency. Since all these impedances contain $j\omega$, we can simplify the writing by dropping this factor altogether and, instead of reasoning in terms of the z-matrix, using the following L-matrix to represent the transformer

$$\begin{Vmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{Vmatrix}. \qquad (329)$$

Since L_{11} and L_{22} are approximately proportional to the squares of the number of primary and secondary turns respectively, the *nominal* ratio of transformation may be defined by

$$a = \sqrt{\frac{L_{22}}{L_{11}}}. \tag{330}$$

The coupling coefficient, according to its usual definition, is

$$k = \frac{L_{12}}{\sqrt{L_{11}L_{22}}}. \tag{331}$$

In any physical case k is less than unity. Zero leakage corresponds to $k = 1$. This would mean

$$L_{11}L_{22} - L_{12}^2 = 0, \tag{332}$$

which is the value of the determinant of the matrix (329). Hence we can say that the vanishing of the determinant of the L-matrix of a transformer is the necessary and sufficient condition for a unity coupling coefficient.

Since the coupling coefficient of the given transformer is not unity, we shall try to split the given network into a series combination of two, of which one shall satisfy the condition for unity coupling. In order to do this we apply the rule for matrix addition and write

$$\begin{Vmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{Vmatrix} = \begin{Vmatrix} L_{11} - L_1 & L_{12} \\ L_{21} & L_{22} - L_2 \end{Vmatrix} + \begin{Vmatrix} L_1 & 0 \\ 0 & L_2 \end{Vmatrix}, \tag{333}$$

and then demand that

$$\begin{vmatrix} L_{11} - L_1 & L_{12} \\ L_{21} & L_{22} - L_2 \end{vmatrix} = (L_{11} - L_1)(L_{22} - L_2) - L_{12}^2 = 0. \tag{334}$$

The first matrix on the right-hand side of (333) represents a transformer without leakage, having the transformation ratio

$$a' = \sqrt{\frac{L_{22} - L_2}{L_{11} - L_1}} = \frac{L_{12}}{L_{11} - L_1} = \frac{L_{22} - L_2}{L_{12}}, \tag{335}$$

while the second matrix represents a T-structure having the inductances L_1 and L_2 in its series arms and a short circuit for its shunt arm. The series combination of these networks gives rise to the circuit of Fig. 41*a* or either of the alternative circuits *b* and *c* of the same figure, involving ideal transformers.

From (334) we can find

$$L_1 = \frac{L_2L_{11} - |L|}{L_2 - L_{22}}; \; L_2 = \frac{L_1L_{22} - |L|}{L_1 - L_{11}}, \tag{336}$$

in which $|L|$ represents the determinant of the matrix (329).

The point of interest lies in the fact that, in these networks which are equivalent to the actual transformer, either L_1 or L_2 can be chosen arbitrarily. Since these inductances represent the leakage of the transformer, this result means that we can distribute the leakage effect between the primary and secondary sides in a way which seems most

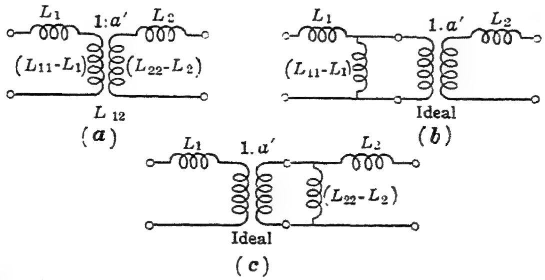


FIG. 41.—Equivalent circuits for the transformer of Fig. 40 showing alternative methods of representing the leakage and winding reactances.

favorable from the standpoint of the network with which the transformer is to be associated. As an illustration we shall consider several special cases.

Suppose we let $L_2 = 0$; then from (331), (335) and (336) we get

$$\left.\begin{aligned} L_1 &= L_{11}(1 - k^2) \\ a' &= \frac{a}{k} \end{aligned}\right\}. \tag{337}$$

The network representations according to Fig. 41 for this case are shown in Fig. 42.

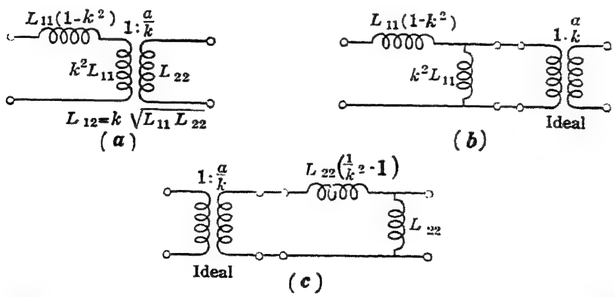


FIG. 42.—Circuits equivalent to those of Figs. 40 and 41 in which all the leakage effect is considered to be on the input side of the transformer.

On the other hand, if we let $L_1 = 0$, then we have

$$\left.\begin{aligned} L_2 &= L_{22}(1 - k^2) \\ a' &= ak \end{aligned}\right\}. \tag{338}$$

The corresponding network representations for this are shown in Fig. 43.

Between these two extremes there are any number of intermediate cases for which the leakage inductance is distributed between the two sides of the transformer. Note, however, that there is nothing hard and fast about the way in which this should be done. This leaves the effective ratio of transformation arbitrary between the limits (337) and (338). In power transformers where the coupling coefficient is very nearly unity, the ratio is for all practical purposes equal to its nominal value. In communication transformers, especially those with an air core, the coupling coefficient may be almost anything between unity and zero, so that the effective ratio becomes quite an arbitrary quantity.

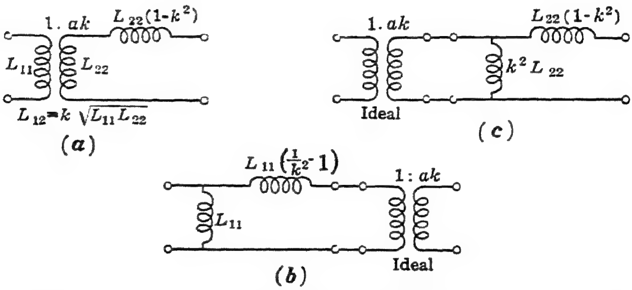


FIG. 43.—Circuits equivalent to those of Figs. 40 and 41 in which all the leakage effect is considered to be on the output side of the transformer.

The results illustrated in Figs. 42 and 43, or any of the intermediate cases, may be useful when the network with which the actual transformer is to be associated possesses inductance in its adjacent series or shunt arms. Under these conditions it may be possible to absorb part or all of the inductances in cases like (*b*) and (*c*), for example, in these adjacent arms. Thus ideal transformer behavior may be approximated or even fully attained except for the presence of winding losses.

7. Discussion of several basic structures. In this section we wish to discuss the y- and z-matrices of several commonly used structures, and from the interrelations of the coefficients of these matrices show how the structures may be made equivalent.

Consider first the T- and Π-structures of Fig. 44, which may be thought of as series and parallel combinations of simpler networks. By inspection we have for their z- and y-systems respectively

$$\left.\begin{array}{l} z_{11} = z_A + z_C;\ z_{22} = z_B + z_C;\ z_{12} = z_C \\ y_{11} = y_A + y_C;\ y_{22} = y_B + y_C;\ y_{12} = -y_C \end{array}\right\}, \tag{339}$$

from which

$$\left.\begin{aligned} z_A &= z_{11} - z_{12};\ z_B = z_{22} - z_{12};\ z_C = z_{12} \\ y_A &= y_{11} + y_{12};\ y_B = y_{22} + y_{12};\ y_C = -y_{12} \end{aligned}\right\}. \tag{340}$$

Noting that

$$\left.\begin{aligned} |z| &= z_A z_B + z_A z_C + z_B z_C \\ |y| &= y_A y_B + y_A y_C + y_B y_C \end{aligned}\right\}, \tag{341}$$

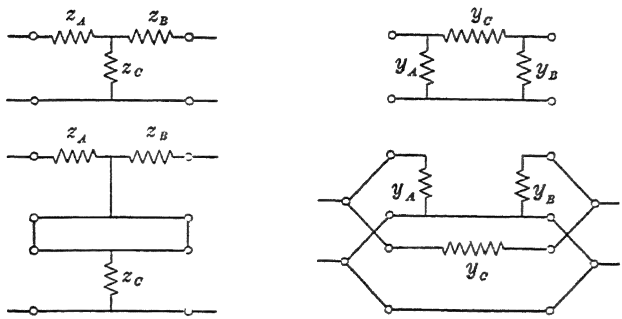


FIG. 44.—Decomposition of T and Π circuits into combinations of simpler networks.

and using the relations (291) and (292) we find for the interrelations between equivalent T- and Π-structures

$$\left.\begin{aligned} z_A &= \frac{y_B}{|y|};\ z_B = \frac{y_A}{|y|};\ z_C = \frac{y_C}{|y|} \\ y_A &= \frac{z_B}{|z|};\ y_B = \frac{z_A}{|z|};\ y_C = \frac{z_C}{|z|} \end{aligned}\right\}, \tag{342}$$

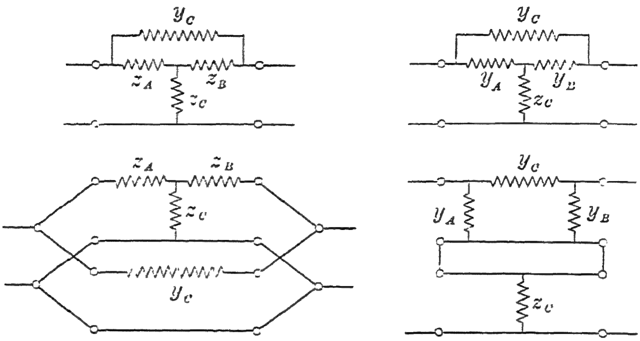
which are the well-known transformations for these structures. Recalling that $|y| = |z|^{-1}$, one set of these transformations can be directly obtained from the other. The simplicity of form is due to the fact that we used admittances to designate the branches in the Π-structure, and impedances for the T-structure.

Another commonly occurring structure is the so-called bridged T which is illustrated in Fig. 45. In part (*a*) of this figure the bridged T is represented as a parallel combination of a T-structure and a Π-structure for which $y_A = y_B = 0$, whereas in part (*b*) it is represented as a series combination of a Π-structure and a T-structure for which $z_A = z_B = 0$. Hence, using (339) and (342), we have for the y-matrix of the bridged T according to (*a*)

$$\left\|\begin{matrix} \dfrac{z_B + z_C}{|z|} & -\dfrac{z_C}{|z|} \\[2ex] -\dfrac{z_C}{|z|} & \dfrac{z_A + z_C}{|z|} \end{matrix}\right\| + \left\|\begin{matrix} y_C & -y_C \\ -y_C & y_C \end{matrix}\right\| = \|y\|, \tag{343}$$

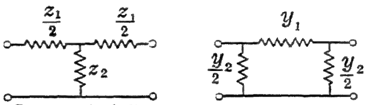
and for the z-matrix according to (b)

$$\left\|\begin{matrix} \dfrac{y_B + y_C}{|y|} & \dfrac{y_C}{|y|} \\[2ex] \dfrac{y_C}{|y|} & \dfrac{y_A + y_C}{|y|} \end{matrix}\right\| + \left\|\begin{matrix} z_C & z_C \\ z_C & z_C \end{matrix}\right\| = \|z\|. \tag{344}$$



(a) (b)

FIG. 45.—Alternative decompositions of the bridged-T structure.



Symmetrical T Symmetrical π

FIG. 46.—Schematic representations of symmetrical T and Π networks.

Symmetrical structures play a particularly important rôle in communication networks. For the T and Π of Fig. 46 we have respectively

$$\|z\| = \left\|\begin{matrix} \dfrac{z_1}{2} + z_2 & z_2 \\[2ex] z_2 & \dfrac{z_1}{2} + z_2 \end{matrix}\right\|, \tag{345}$$

and

$$\|y\| = \left\|\begin{matrix} \dfrac{y_2}{2} + y_1 & -y_1 \\[2ex] -y_1 & \dfrac{y_2}{2} + y_1 \end{matrix}\right\|. \tag{346}$$

The symmetrical bridged T may be obtained by combining these after the fashion of Fig. 45.

Another common symmetrical structure is the **lattice or bridge.** As shown in Fig. 47, this may be thought of as a parallel combination of two simpler structures. Writing the y-matrices for the upper and

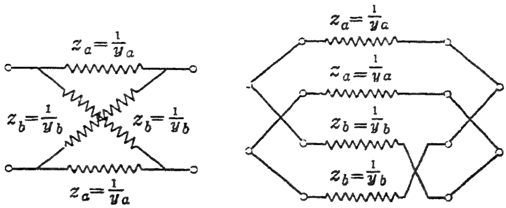


FIG. 47.—The symmetrical lattice network and its representation as a parallel combination of simpler ones.

lower of these simpler structures by inspection and adding we have for the y-matrix of the lattice

$$\begin{Vmatrix} \frac{y_a}{2} & -\frac{y_a}{2} \\ -\frac{y_a}{2} & \frac{y_a}{2} \end{Vmatrix} + \begin{Vmatrix} \frac{y_b}{2} & \frac{y_b}{2} \\ \frac{y_b}{2} & \frac{y_b}{2} \end{Vmatrix} = \begin{Vmatrix} \frac{y_b + y_a}{2} & \frac{y_b - y_a}{2} \\ \frac{y_b - y_a}{2} & \frac{y_b + y_a}{2} \end{Vmatrix} \tag{347}$$

Converting by means of (292) we have for the z-matrix of the symmetrical lattice

$$\begin{Vmatrix} \frac{z_b + z_a}{2} & \frac{z_b - z_a}{2} \\ \frac{z_b - z_a}{2} & \frac{z_b + z_a}{2} \end{Vmatrix} \tag{348}$$

From these we may easily obtain for the component impedances or admittances of the symmetrical lattice

$$\left.\begin{aligned} z_a &= z_{11} - z_{12};\ z_b = z_{11} + z_{12} \\ y_a &= y_{11} - y_{12};\ y_b = y_{11} + y_{12} \end{aligned}\right\} \tag{349}$$

In order to determine the equivalent symmetrical T- or Π-structures for the lattice, we have merely to equate the matrices (345) and (348) or (346) and (347) respectively, thus

$$\begin{Vmatrix} \frac{z_1}{2} + z_2 & z_2 \\ z_2 & \frac{z_1}{2} + z_2 \end{Vmatrix} = \begin{Vmatrix} \frac{z_b + z_a}{2} & \frac{z_b - z_a}{2} \\ \frac{z_b - z_a}{2} & \frac{z_b + z_a}{2} \end{Vmatrix} \tag{350}$$

and

$$\begin{Vmatrix} \frac{y_2}{2} + y_1 & -y_1 \\ -y_1 & \frac{y_2}{2} + y_1 \end{Vmatrix} = \begin{Vmatrix} \frac{y_b + y_a}{2} & \frac{y_b - y_a}{2} \\ \frac{y_b - y_a}{2} & \frac{y_b + y_a}{2} \end{Vmatrix} \tag{351}$$

Since two matrices are equal only when all corresponding coefficients are equal, we can write the equivalence equations for these cases by inspection. Thus the lattice-to-T or-Π conversions become

$$\left.\begin{aligned} z_1 &= 2z_a; \qquad z_2 = \frac{z_b - z_a}{2} \\ y_1 &= \frac{y_a - y_b}{2}; \; y_2 = 2y_b \end{aligned}\right\}, \tag{352}$$

while for the T- or Π-to-lattice conversions we have

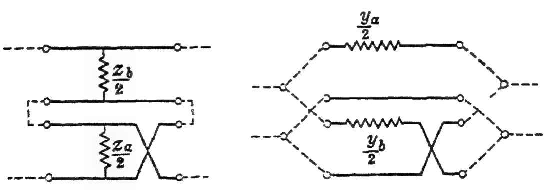
$$\left.\begin{aligned} z_b &= \frac{z_1}{2} + 2z_2; \; z_a = \frac{z_1}{2} \\ y_a &= \frac{y_2}{2} + 2y_1; \; y_b = \frac{y_2}{2} \end{aligned}\right\}. \tag{353}$$

Note that, whereas the latter conversions always lead to a physically realizable lattice, the conversions (352) may not lead to a physically realizable T or Π owing to the appearance of negative terms in these equations.

There are several other structures equivalent to the lattice which will be useful to us later on. These are obtained from the matrix expressions (347) and (348). For the latter matrix we can write

$$\begin{Vmatrix} \frac{z_b}{2} & \frac{z_b}{2} \\ \frac{z_b}{2} & \frac{z_b}{2} \end{Vmatrix} + \begin{Vmatrix} \frac{z_a}{2} & -\frac{z_a}{2} \\ -\frac{z_a}{2} & \frac{z_a}{2} \end{Vmatrix}. \tag{348a}$$

This and the left-hand side of (347) suggest that equivalent lattice structures will be obtained by combining in series or parallel the net-



(a) (b)

FIG. 48.—Symmetrical lattice equivalents suggested by the decomposition illustrated in Fig. 47, but for which the interconnection tests in Fig. 36 are not satisfied.

works represented by the individual matrices, as indicated in Fig. 48. These are cases, however, for which the matrix combinations are invalid unless ideal transformers are used. When this is done on the output sides of the lower networks, the resulting equivalents take the forms

shown in Fig. 49 in which the ideal transformers are indicated as having a $-$ 1:1 ratio. This is equivalent to having the output terminals of the lower networks in Fig. 48 crossed.

A further modification based upon the relation (347) for the y-matrix leads to a third alternative structure. Fig. 50 shows the derivation of this form. The upper network in part (a) of this figure represents the first matrix in (347); the lower one represents the second matrix. This is easily checked by inspection. Part (b) of the figure, which is obtained from part (a) by redrawing it and combining the two admittances y_b in parallel, is the desired alternative.

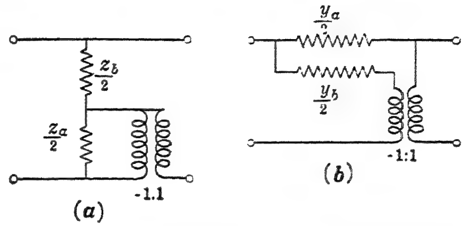


FIG. 49.—Symmetrical lattice equivalents according to the scheme illustrated in Fig. 48 for which the validity conditions for interconnection are satisfied by the ideal transformer method in Fig. 37.

These networks, which are equivalent to the lattice, are more economical in that they contain only two branches whereas the lattice requires four. Unless the ideal transformers can be sufficiently approximated, however, these structures are of little practical value.

8. Characteristic impedance and propagation functions. In the preceding chapter we saw that the external behavior of the long trans-

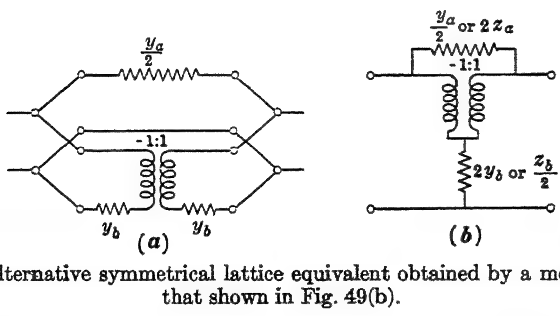


FIG. 50.—Alternative symmetrical lattice equivalent obtained by a modification of that shown in Fig. 49(b).

mission line could be completely specified by means of its sending- and receiving-end voltages and currents, and that their interrelations in turn were determined in terms of two functions, namely the characteristic impedance and propagation functions. Although the general type of four-terminal network which we are here considering need not be the analytic representation of a long line, it seems desirable that we should

attempt to specify its behavior in similar terms. This way of expressing four-terminal network behavior in general may prove particularly useful if several sections of lines or cables are cascaded or otherwise associated with various lumped-constant networks. To express the behavior of the latter in terms similar to those which are used for the line sections will tend to unify and perhaps simplify the treatment of the combined system.

The most logical way in which to develop a treatment for four-terminal network behavior which will be entirely analogous to that used in connection with the long line is to consider a chain or cascade

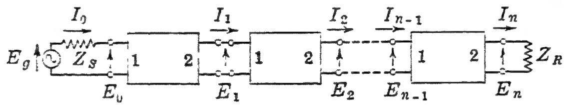


FIG. 51.—Cascade of identical dissymmetrical networks on the iterative basis.

of identical networks working out of and into impedances Z_S and Z_R respectively and driven harmonically at the sending end. Such a system is called a uniform recurrent structure. One form is illustrated in Fig. 51. In this arrangement the networks are all oriented in the same direction; i.e., ends 1 are on the left and ends 2 on the right. An important modification of this scheme will be discussed later.

The voltages and currents at the junctions are numbered from E_0 and I_0 to E_n and I_n, n being the number of networks in the chain. Given the four-terminal network characteristics, the generator voltage E_g, and the impedances Z_S and Z_R, the problem is to find analytic expressions for the voltage and current variations along the structure, i.e., from junction to junction. The physical aspects of this problem are very similar to those which we found in the long line. The major difference lies in the fact that the voltages and currents here are discontinuous functions of distance along the structure, the various junctions being referred to by subscripts.

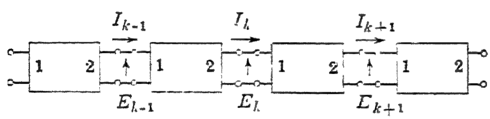


FIG. 52.—Internal portion of the cascaded structure of Fig. 51.

The method of analysis which we shall apply is entirely analogous to that used in solving the long-line problem. We shall first write the voltage and current equilibrium equations for a typical internal section of the structure, and, having found formal solutions to these, proceed to evaluate the various arbitrary constants involved by means of the boundary conditions.

In Fig. 52 is shown an internal portion of the uniform structure.

Applying the y-system (275) to the junction k, first considering I_k as the output of the preceding network and then as the input to the succeeding one, we have with due attention to the current directions

$$I_k = -y_{21}E_{k-1} - y_{22}E_k = y_{11}E_k + y_{12}E_{k+1}, \tag{354}$$

from which

$$E_{k-1} + \left(\frac{y_{11} + y_{22}}{y_{12}}\right)E_k + E_{k+1} = 0. \tag{355}$$

Applying the z-system (276) to the junction k in the same manner we get

$$E_k = z_{21}I_{k-1} - z_{22}I_k = z_{11}I_k - z_{12}I_{k+1}, \tag{356}$$

so that

$$I_{k-1} - \left(\frac{z_{11} + z_{22}}{z_{12}}\right)I_k + I_{k+1} = 0. \tag{357}$$

From (291) and (292) we see that

$$\frac{y_{11} + y_{22}}{y_{12}} = -\frac{z_{11} + z_{22}}{z_{12}} = -(\mathfrak{A} + \mathfrak{D}). \tag{358}$$

Hence the voltage and current equilibrium conditions (355) and (357) become

$$\left.\begin{aligned} E_{k-1} - (\mathfrak{A} + \mathfrak{D})E_k + E_{k+1} &= 0 \\ I_{k-1} - (\mathfrak{A} + \mathfrak{D})I_k + I_{k+1} &= 0 \end{aligned}\right\}. \tag{359}$$

These are analogous to the equations (42) of Chapter II which express the equilibrium of a typical internal section of the uniform line. As solutions we assumed the exponential forms (45) and (46). Heuristically we can, therefore, attempt a solution of (359) by assuming similar forms, namely

$$\left.\begin{aligned} E_k &= A_1e^{-k\gamma} + A_2e^{k\gamma} \\ I_k &= B_1e^{-k\gamma} + B_2e^{k\gamma} \end{aligned}\right\}, \tag{360}$$

where γ plays the same part for the uniform recurrent structure that α plays for the uniform line, and the index k takes the place of the continuous variable x. Thus γ becomes the propagation function per network (also called section) in the recurrent structure just as α is the propagation function per unit length in the line.

Substituting our assumed solutions (360) into (359) we find after some factoring that the former will be valid subject to the conditions

$$\left.\begin{aligned} [e^{\gamma} - (\mathfrak{A} + \mathfrak{D}) + e^{-\gamma}]E_k &= 0 \\ [e^{\gamma} - (\mathfrak{A} + \mathfrak{D}) + e^{-\gamma}]I_k &= 0 \end{aligned}\right\}. \tag{361}$$

Non-vanishing solutions of the form (360), therefore, are obtained provided

$$e^{\gamma} + e^{-\gamma} = (\mathfrak{A} + \mathfrak{D}), \tag{362}$$

which determines the propagation function γ. Namely, (362) gives

$$\cosh \gamma = \frac{\mathfrak{A} + \mathfrak{D}}{2}, \tag{363}$$

so that

$$\sinh \gamma = \frac{\sqrt{(\mathfrak{A} + \mathfrak{D})^2 - 4}}{2}, \tag{364}$$

and

$$e^{\pm\gamma} = \frac{(\mathfrak{A} + \mathfrak{D}) \pm \sqrt{(\mathfrak{A} + \mathfrak{D})^2 - 4}}{2}, \tag{365}$$

from which

$$\gamma = \ln\left(\frac{(\mathfrak{A} + \mathfrak{D}) + \sqrt{(\mathfrak{A} + \mathfrak{D})^2 - 4}}{2}\right). \tag{366}$$

Returning to the solutions (360), the next step is to evaluate the constants A_1, A_2, B_1, and B_2. Here we are reminded that the interrelation between voltage and current as expressed by (354) or (356) will suffice to determine two of these constants in terms of the other two. Since (354) and (356) are not independent, either one may be used. Thus substituting (360) into (356) we have

$$A_1e^{-k\gamma} + A_2e^{k\gamma} = z_{11}B_1e^{-k\gamma} + z_{11}B_2e^{k\gamma} - z_{12}B_1e^{-(k+1)\gamma} - z_{12}B_2e^{(k+1)\gamma}. \tag{367}$$

Equating coefficients of $e^{k\gamma}$ and $e^{-k\gamma}$ gives

$$\left.\begin{aligned} A_1 &= (z_{11} - z_{12}e^{-\gamma})B_1 \\ A_2 &= (z_{11} - z_{12}e^{\gamma})B_2 \end{aligned}\right\}. \tag{368}$$

Although it is rather artificial to speak of the behavior of a recurrent lumped structure in terms of wave propagation, the analysis of such a system from the standpoint of its characteristic impedance and propagation function suggests that the above solution be interpreted in this sense. Thus A_1 and A_2 may be designated as the incident and reflected amplitudes of the net voltage, and B_1 and B_2 as the corresponding current amplitudes. Recalling the equations (54) which were found to relate these quantities in the long line, the analogy between that case and our present one suggests that we regard the factors of B_1 and B_2 in (368) as the characteristic impedances, with positive and negative sign respectively, of the recurrent lumped structure. Since the general case which we are treating is dissymmetrical, it is logical to expect that the characteristic impedances should be different in the

positive and negative directions, i.e., for the incident and the reflected wave amplitudes. Let us denote these by Z_{01} and Z_{02} respectively, and write

$$\left.\begin{aligned} B_1 &= \frac{A_1}{Z_{01}} \\ B_2 &= \frac{A_2}{-Z_{02}} \end{aligned}\right\}. \tag{369}$$

Then comparing with (368) and using (365) we find

$$\left.\begin{aligned} Z_{01} &= \frac{\sqrt{(\mathfrak{A}+\mathfrak{D})^2-4}+(\mathfrak{A}-\mathfrak{D})}{2\mathfrak{C}} \\ Z_{02} &= \frac{\sqrt{(\mathfrak{A}+\mathfrak{D})^2-4}-(\mathfrak{A}-\mathfrak{D})}{2\mathfrak{C}} \end{aligned}\right\}. \tag{370}$$

At the boundaries of the structure we may write the following condition equations

$$\left.\begin{aligned} E_0 + I_0 Z_S &= E_g \\ E_n - I_n Z_R &= 0 \end{aligned}\right\}. \tag{371}$$

Substituting (360) and (369) into these we get

$$\left.\begin{aligned} \left(\frac{Z_S+Z_{01}}{Z_{01}}\right)A_1 \qquad - \left(\frac{Z_S-Z_{02}}{Z_{02}}\right)A_2 &= E_g \\ -\left(\frac{Z_R-Z_{01}}{Z_{01}}\right)e^{-n\gamma}A_1 + \left(\frac{Z_R+Z_{02}}{Z_{02}}\right)e^{n\gamma}A_2 &= 0 \end{aligned}\right\}. \tag{372}$$

Denoting the determinant of this system by Δ we have

$$\Delta = \frac{(Z_S+Z_{01})(Z_R+Z_{02})e^{n\gamma} - (Z_S-Z_{02})(Z_R-Z_{01})e^{-n\gamma}}{Z_{01}Z_{02}}, \tag{373}$$

and

$$A_1 = \frac{(Z_R+Z_{02})e^{n\gamma}E_g}{Z_{02}\cdot\Delta}; \quad A_2 = \frac{(Z_R-Z_{01})e^{-n\gamma}E_g}{Z_{01}\cdot\Delta}. \tag{374}$$

Defining as the reflection coefficients

$$r_S = \frac{Z_S-Z_{02}}{Z_S+Z_{01}}; \quad r_R = \frac{Z_R-Z_{01}}{Z_R+Z_{02}}, \tag{375}$$

the voltage and current solutions take the final form

$$\left.\begin{aligned} E_k &= \frac{E_g(Z_{01}e^{(n-k)\gamma} + Z_{02}r_R e^{-(n-k)\gamma})}{(Z_S+Z_{01})(e^{n\gamma} - r_S r_R e^{-n\gamma})} \\ I_k &= \frac{E_g(e^{(n-k)\gamma} - r_R e^{-(n-k)\gamma})}{(Z_S+Z_{01})(e^{n\gamma} - r_S r_R e^{-n\gamma})} \end{aligned}\right\}. \tag{376}$$

These results may, of course, be put into any of the alternative forms discussed in Chapter II in connection with the long-line problem. The various voltage and current ratios given there apply here also. In fact, the further discussion of the results for this case is so similar to that already given for the long line that we shall not consume further space in elaborating upon the solutions to our present problem. One feature, however, is worth pointing out. Namely, if we consider the input voltage and current obtained by putting $k = 0$, and then let n approach infinity, the exponential terms with $e^{-n\gamma}$ will vanish, leaving

$$E_0 = \frac{Z_{01}E_g}{Z_S + Z_{01}}; \; I_0 = \frac{E_g}{Z_S + Z_{01}}, \tag{377}$$

so that we have for the ratio of these

$$\left(\frac{E_0}{I_0}\right)_{n \to \infty} = Z_{01}, \tag{378}$$

which gives a physical interpretation to the characteristic impedance Z_{01}. Namely, it is that impedance looking into an infinitely long chain of identical dissymmetrical structures all of which are oriented with their ends 1 on the left and ends 2 on the right, i.e., as shown in Fig. 51. The same input impedance is obtained with a finite structure terminated in Z_{01}, which makes $r_R = 0$.

If the subscripts 1 in these solutions are interchanged with 2, and the parameter $\mathfrak{A}$ interchanged with $\mathfrak{D}$, we have the corresponding solution for the structure with all networks reversed. Then (378) will define Z_{02}. These impedances are also referred to as the **iterative impedances** of the structure; and a termination in that iterative impedance which takes the place of an infinite continuation is called **termination on an iterative basis.** This basis of termination, however, does not make the transmission from source to load of the volt-ampere product a maximum, and it is, therefore, not the one commonly used in communication work. The reason for this can at least be roughly appreciated from the fact that termination on an iterative basis does not lead to impedance symmetry at the junctions unless the networks happen to be symmetrical. By impedance symmetry we understand that condition at a junction for which the impedance looking in both directions at that junction is the same. In the above example, if $Z_S = Z_{02}$ and $Z_R = Z_{01}$, then the impedance looking to the left at any junction is evidently Z_{02} while looking to the right we see Z_{01}. This dissymmetry at the junctions may be obviated by reorienting the networks in the uniform structure. This we shall discuss in more detail now.

Instead of the structure of Fig. 51, consider the cascade arrangement shown in Fig. 53. Here the networks are alternately reversed so that similar ends are everywhere adjacent. The junctions are, therefore, symmetrical although the networks are themselves dissymmetrical. A complication enters into the analysis of this case, however, owing to the fact that we have here two kinds of junctions. Hence we have no right to expect that the same functional relation for voltage or current will hold for the odd as well as for the even subscripts. According to Fig. 53 the even subscripts refer to the junctions of ends 1 and the odd

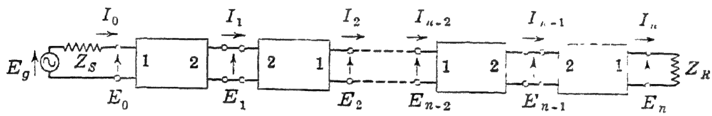


FIG. 53.—Cascade of identical dissymmetrical networks on the image basis

subscripts refer to the junctions of ends 2. Furthermore, the final end to which Z_R is connected will be an end 1 when n is even and an end 2 when n is odd. Hence a distinction will have to be made between these cases. The two additional cases which are obtained from these respectively by reversing all networks obviously need not be given special treatment since they can be obtained from the others by simply interchanging subscripts 1 and 2 and parameters $\mathfrak{A}$ and $\mathfrak{D}$. We shall, therefore, consider only the structure of Fig. 53 for the cases n *even* and n *odd*.

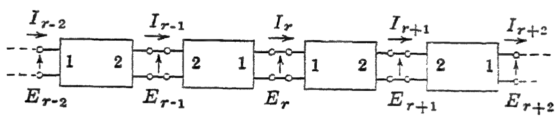


FIG. 54.—Internal portion of the cascaded structure of Fig. 53.

We begin again by writing the voltage and current equilibrium conditions for a typical internal portion of the structure. This is illustrated in Fig. 54. Comparison with Fig. 53 shows that the subscript r must be an *even* integer. Applying the fundamental relations (275) successively to the junctions $r - 2$ to $r + 1$, with due regard to the orientation of the networks and directions of the currents, and considering the currents first as the output from the preceding network and then as the input to the succeeding one, we get the following series of equations

$$\left.\begin{aligned} I_{r-2} &= -y_{12}E_{r-3} - y_{11}E_{r-2} = y_{11}E_{r-2} + y_{12}E_{r-1} \\ I_{r-1} &= -y_{21}E_{r-2} - y_{22}E_{r-1} = y_{22}E_{r-1} + y_{21}E_r \\ I_r &= -y_{12}E_{r-1} - y_{11}E_r \quad = y_{11}E_r \quad + y_{12}E_{r+1} \\ I_{r+1} &= -y_{21}E_r \quad - y_{22}E_{r+1} = y_{22}E_{r+1} + y_{21}E_{r+2} \end{aligned}\right\}. \tag{379}$$

Using the right-hand sides of these, and substituting for the y's by means of (291), we get

$$\left.\begin{aligned} &E_{r-3} - 2\mathfrak{D}E_{r-2} + E_{r-1} = 0 \\ &\qquad E_{r-2} - 2\mathfrak{A}E_{r-1} + E_r = 0 \\ &\qquad\qquad E_{r-1} - 2\mathfrak{D}E_r + E_{r+1} = 0 \\ &\qquad\qquad\qquad E_r - 2\mathfrak{A}E_{r+1} + E_{r+2} = 0 \end{aligned}\right\}. \tag{380}$$

Applying the relations (276) in a similar manner we have

$$\left.\begin{aligned} E_{r-2} &= z_{12}I_{r-3} - z_{11}I_{r-2} = z_{11}I_{r-2} - z_{12}I_{r-1} \\ E_{r-1} &= z_{21}I_{r-2} - z_{22}I_{r-1} = z_{22}I_{r-1} - z_{21}I_r \\ E_r &= z_{12}I_{r-1} - z_{11}I_r = z_{11}I_r - z_{12}I_{r+1} \\ E_{r+1} &= z_{21}I_r - z_{22}I_{r+1} = z_{22}I_{r+1} - z_{21}I_{r+2} \end{aligned}\right\}. \tag{381}$$

With these and the substitutions (292) we have

$$\left.\begin{aligned} &I_{r-3} - 2\mathfrak{A}I_{r-2} + I_{r-1} = 0 \\ &\qquad I_{r-2} - 2\mathfrak{D}I_{r-1} + I_r = 0 \\ &\qquad\qquad I_{r-1} - 2\mathfrak{A}I_r + I_{r+1} = 0 \\ &\qquad\qquad\qquad I_r - 2\mathfrak{D}I_{r+1} + I_{r+2} = 0 \end{aligned}\right\}. \tag{382}$$

The fact that the equations (380) and (382) alternately contain the parameters $\mathfrak{A}$ and $\mathfrak{D}$ shows that the voltage and current relations do not follow a uniform variation from junction to junction. This is to be expected since the junctions are alternately those of ends 1 and 2. However, if we derive relations for the voltage and current which skip every other junction, i.e., progress by two networks at one time, then we may expect a uniform result.

If in (380) we multiply the second equation by $2\mathfrak{D}$, and then add the first three; or multiply the third equation by $2\mathfrak{A}$ and add the last three we get respectively

$$\left.\begin{aligned} E_{r-3} - 2(2\mathfrak{A}\mathfrak{D} - 1)E_{r-1} + E_{r+1} = 0 \\ E_{r-2} - 2(2\mathfrak{A}\mathfrak{D} - 1)E_r + E_{r+2} = 0 \end{aligned}\right\}. \tag{383}$$

Treating the current equations (382) in a similar manner we have

$$\left.\begin{aligned} I_{r-3} - 2(2\mathfrak{A}\mathfrak{D} - 1)I_{r-1} + I_{r+1} = 0 \\ I_{r-2} - 2(2\mathfrak{A}\mathfrak{D} - 1)I_r + I_{r+2} = 0 \end{aligned}\right\}. \tag{384}$$

These results show that a uniform relation can be expected for the variation of either voltage or current along alternate junctions. Since these equations all contain the same parameter $2(2\mathfrak{A}\mathfrak{D} - 1)$, they also show that the propagation properties along the odd or even junctions

are the same, and furthermore are the same for voltage and current. This makes it possible to assume the following forms

$$\left.\begin{array}{ll} E_k = A_1 e^{-k\theta} + A_2 e^{k\theta} & \text{for } k \text{ even} \\ E_k = A_1' e^{-k\theta} + A_2' e^{k\theta} & \text{for } k \text{ odd} \end{array}\right\}, \tag{385}$$

and

$$\left.\begin{array}{ll} I_k = B_1 e^{-k\theta} + B_2 e^{k\theta} & \text{for } k \text{ even} \\ I_k = B_1' e^{-k\theta} + B_2' e^{k\theta} & \text{for } k \text{ odd} \end{array}\right\}, \tag{386}$$

i.e., the solutions for the odd and even junctions can differ only by their amplitudes.

For the determination of the propagation function, which is here denoted by θ, we may substitute either of the assumed solutions (385) or (386) into (383) or (384). Suppose we substitute the first equation (385) into the second equation (383). After factoring, this gives

$$(A_1 e^{-k\theta} + A_2 e^{k\theta})\{e^{2\theta} - 2(2\mathfrak{A}\mathfrak{D} - 1) + e^{-2\theta}\} = 0.$$

A non-vanishing solution, therefore, requires

$$2\mathfrak{A}\mathfrak{D} - 1 = \cosh 2\theta,$$

or

$$\mathfrak{A}\mathfrak{D} = \frac{1 + \cosh 2\theta}{2} = \cosh^2\theta,$$

so that

$$\left.\begin{array}{r} \cosh\theta = \sqrt{\mathfrak{A}\mathfrak{D}} \\ \sinh\theta = \sqrt{\mathfrak{A}\mathfrak{D} - 1} = \sqrt{\mathfrak{B}\mathfrak{C}} \\ e^{\pm\theta} = (\sqrt{\mathfrak{A}\mathfrak{D}} \pm \sqrt{\mathfrak{B}\mathfrak{C}}) \end{array}\right\}. \tag{387}$$

For the determination of A_1' and A_2' in terms of A_1 and A_2, we substitute (385) into any of the equations (380) which relate successive junctions. Here we must remember that r was taken to designate an *even* integer. Hence if we choose the third equation (380), we must use the second solution (385) for the first and last terms, and the first solution (385) for the second term. Thus we get after some factoring

$$A_1' e^{-r\theta}\left(e^{\theta} - 2\mathfrak{D}\frac{A_1}{A_1'} + e^{-\theta}\right) + A_2' e^{r\theta}\left(e^{-\theta} - 2\mathfrak{D}\frac{A_2}{A_2'} + e^{\theta}\right) = 0, \tag{388}$$

which must hold for all even values of r. This requires that the coefficients of $e^{r\theta}$ and $e^{-r\theta}$ vanish separately. For non-vanishing values of A_1' and A_2' this requires

$$e^{\theta} - 2\mathfrak{D}\frac{A_1}{A_1'} + e^{-\theta} = 0,$$

and

$$e^{-\theta} - 2\mathfrak{D}\frac{A_2}{A_2'} + e^{\theta} = 0.$$

Making use of the relations (387) this gives

$$\frac{A_1}{A_1'} = \frac{A_2}{A_2'} = \sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}} \tag{389}$$

which is the desired relationship.

Similarly, if we substitute the solutions (386) into the third equation (382), bearing in mind that r is an *even* integer, and demand non-vanishing values for B_1' and B_2', we get

$$e^{\theta} - 2\mathfrak{A}\frac{B_1}{B_1'} + e^{-\theta} = 0,$$

and

$$e^{-\theta} - 2\mathfrak{A}\frac{B_2}{B_2'} + e^{\theta} = 0,$$

from which we find

$$\frac{B_1}{B_1'} = \frac{B_2}{B_2'} = \sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}}. \tag{390}$$

The results (389) and (390) express the solutions for odd integers in (385) and (386) in terms of those for even integers.

The relations between the voltage and current amplitudes are found by substituting these solutions into any of the left-hand equations (379) or (381). For example, substituting into the third equation (381) gives

$$A_1e^{-k\theta} + A_2e^{k\theta} = z_{12}\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}(B_1e^{-(k-1)\theta} + B_2e^{(k-1)\theta}) - z_{11}(B_1e^{-k\theta} + B_2e^{k\theta}).$$

Equating coefficients of $e^{-k\theta}$ and $e^{k\theta}$ we have

$$A_1 = \left(z_{12}\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}e^{\theta} - z_{11}\right)B_1$$

and

$$A_2 = \left(z_{12}\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}e^{-\theta} - z_{11}\right)B_2.$$

Substituting for the z's from (292) and using the last relation (387) this gives

$$\left.\begin{aligned} A_1 &= \sqrt{\frac{\mathfrak{A}\mathfrak{B}}{\mathfrak{C}\mathfrak{D}}} \cdot B_1 \\ A_2 &= -\sqrt{\frac{\mathfrak{A}\mathfrak{B}}{\mathfrak{C}\mathfrak{D}}} \cdot B_2 \end{aligned}\right\}. \tag{391}$$

The factor of B_1 or minus the factor of B_2 in these relations must be interpreted as the characteristic impedance of the recurrent structure

for this case. Furthermore, since we are considering the structure as beginning with an end 1, we must distinguish the characteristic impedance by a suitable subscript. We shall write it as

$$Z_{I1} = \sqrt{\frac{\mathfrak{A}\mathfrak{B}}{\mathfrak{C}\mathfrak{D}}}. \tag{392}$$

The I in the subscript is added to distinguish this impedance from the iterative impedances.

With (389), (390), (391), and (392) we may now rewrite our solutions in the form

$$\left.\begin{aligned} E_k &= A_1 e^{-k\theta} + A_2 e^{k\theta} && \text{for } k \text{ even} \\ E_k &= \sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}}\,(A_1 e^{-k\theta} + A_2 e^{k\theta}) && \text{for } k \text{ odd} \end{aligned}\right\}. \tag{393}$$

$$\left.\begin{aligned} I_k &= \frac{A_1}{Z_{I1}} e^{-k\theta} - \frac{A_2}{Z_{I1}} e^{k\theta} && \text{for } k \text{ even} \\ I_k &= \sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}\left(\frac{A_1}{Z_{I1}} e^{-k\theta} - \frac{A_2}{Z_{I1}} e^{k\theta}\right) && \text{for } k \text{ odd} \end{aligned}\right\}. \tag{394}$$

The last step is to evaluate A_1 and A_2 from boundary conditions. These are the same as for the preceding case and are, therefore, given by (371). In substituting into these we must, however, distinguish between *n odd* and *n even* in writing the condition for the end of the structure. The first boundary condition (371) gives

$$(Z_S + Z_{I1})A_1 - (Z_S - Z_{I1})A_2 = Z_{I1}E_g, \tag{395}$$

while the second boundary condition gives

$$-(Z_R - Z_{I1})e^{-n\theta}A_1 + (Z_R + Z_{I1})e^{n\theta}A_2 = 0 \tag{396}$$

for *n even*, and

$$-(Z_R - Z_{I2})e^{-n\theta}A_1 + (Z_R + Z_{I2})e^{n\theta}A_2 = 0 \tag{397}$$

for *n odd*. In the last equation we have introduced

$$Z_{I2} = \frac{\mathfrak{D}}{\mathfrak{A}} \cdot Z_{I1} = \sqrt{\frac{\mathfrak{B}\mathfrak{D}}{\mathfrak{A}\mathfrak{C}}}, \tag{398}$$

which is the same as Z_{I1} except that $\mathfrak{A}$ and $\mathfrak{D}$ are interchanged.

Solving (395) with either (396) or (397), and defining the reflection coefficients,

$$\left.\begin{aligned} r_{S1} &= \frac{Z_S - Z_{I1}}{Z_S + Z_{I1}}; & r_{R1} &= \frac{Z_R - Z_{I1}}{Z_R + Z_{I1}} \\ r_{S2} &= \frac{Z_S - Z_{I2}}{Z_S + Z_{I2}}; & r_{R2} &= \frac{Z_R - Z_{I2}}{Z_R + Z_{I2}} \end{aligned}\right\}, \tag{399}$$

the solutions (393) and (394) become for n *even*

$$\left.\begin{aligned} E_k &= \frac{Z_{I1}E_g(e^{(n-k)\theta} + r_{R1}e^{-(n-k)\theta})}{(Z_S + Z_{I1})(e^{n\theta} - r_{S1}r_{R1}e^{-n\theta})} && \text{for } k \text{ even} \\ E_k &= \sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}}\left\{ \quad " \quad " \quad " \quad \right\} && \text{for } k \text{ odd} \end{aligned}\right\}. \tag{400}$$

$$\left.\begin{aligned} I_k &= \frac{E_g(e^{(n-k)\theta} - r_{R1}e^{-(n-k)\theta})}{(Z_S + Z_{I1})(e^{n\theta} - r_{S1}r_{R1}e^{-n\theta})} && \text{for } k \text{ even} \\ I_k &= \sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}\left\{ \quad " \quad " \quad " \quad \right\} && \text{for } k \text{ odd} \end{aligned}\right\}. \tag{401}$$

and for n *odd*

$$\left.\begin{aligned} E_k &= \frac{Z_{I1}E_g(e^{(n-k)\theta} + r_{R2}e^{-(n-k)\theta})}{(Z_S + Z_{I1})(e^{n\theta} - r_{S1}r_{R2}e^{-n\theta})} && \text{for } k \text{ even} \\ E_k &= \sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}}\left\{ \quad " \quad " \quad " \quad \right\} && \text{for } k \text{ odd} \end{aligned}\right\}. \tag{402}$$

and

$$\left.\begin{aligned} I_k &= \frac{E_g(e^{(n-k)\theta} - r_{R2}e^{-(n-k)\theta})}{(Z_S + Z_{I1})(e^{n\theta} - r_{S1}r_{R2}e^{-n\theta})} && \text{for } k \text{ even} \\ I_k &= \sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}\left\{ \quad " \quad " \quad " \quad \right\} && \text{for } k \text{ odd} \end{aligned}\right\}. \tag{403}$$

These results may likewise be put into any of the alternative forms discussed in connection with the long-line problem. From them we may form any of the voltage, current, or volt-ampere ratios for arbitrary terminations. If, in particular, the structure is either infinitely long or finite and terminated in Z_{I1} for n even or in Z_{I2} for n odd so that r_{R1} or r_{R2} becomes zero, we have for n either even or odd

$$\left.\begin{aligned} E_k &= \frac{Z_{I1}E_g e^{-k\theta}}{Z_S + Z_{I1}} && \text{for } k \text{ even} \\ E_k &= \sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}}\left(\quad " \quad \right) && \text{for } k \text{ odd} \end{aligned}\right\}, \tag{404}$$

and

$$\left.\begin{aligned} I_k &= \frac{E_g e^{-k\theta}}{Z_S + Z_{I1}} && \text{for } k \text{ even} \\ I_k &= \sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}\left(\quad " \quad \right) && \text{for } k \text{ odd} \end{aligned}\right\}. \tag{405}$$

From these we get for $k = 0$

$$\frac{E_0}{I_0} = Z_{I1}, \tag{406}$$

which gives Z_{I1} the significance of being that impedance looking into

an infinite structure of this type starting with an end 1. Likewise the impedance looking into an infinite structure starting with an end 2 is Z_{I2}. When a finite structure of this type is terminated at either end in its proper impedance Z_{I1} or Z_{I2}, then at any junction the same impedance will be seen in either direction. Hence these characteristic impedances are also called the **image impedances** of the four-terminal network. A network terminated in its proper image impedances is said to be **terminated on an image basis.** The propagation function θ, given by (387), may be referred to as the **image propagation function** to distinguish it from γ which was obtained for the iterative structure. Some engineers also refer to θ as the **image transfer constant.**

If we consider the structure terminated in its proper image impedance at either end, the voltage and current variations along it become very simple. From (404) we get for $Z_S = Z_{I1}$

$$\left.\begin{aligned} E_k &= \frac{E_g}{2} e^{-k\theta} && \text{for } k \text{ even} \\ E_k &= \sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}} \cdot \frac{E_g}{2} \cdot e^{-k\theta} && \text{for } k \text{ odd} \end{aligned}\right\}, \tag{404a}$$

and from (405)

$$\left.\begin{aligned} I_k &= \frac{E_g}{2Z_{I1}} e^{-k\theta} && \text{for } k \text{ even} \\ I_k &= \sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}} \cdot \frac{E_g}{2Z_{I1}} e^{-k\theta} && \text{for } k \text{ odd} \end{aligned}\right\}. \tag{405a}$$

The ratios of the input voltage or current to those at any other junction are then

$$\left(\frac{E_0}{E_k}\right)_{\text{even}} = e^{k\theta}; \left(\frac{E_0}{E_k}\right)_{\text{odd}} = \sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}} \cdot e^{k\theta}, \tag{407}$$

and

$$\left(\frac{I_0}{I_k}\right)_{\text{even}} = e^{k\theta}; \left(\frac{I_0}{I_k}\right)_{\text{odd}} = \sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}} \cdot e^{k\theta}. \tag{407a}$$

It is interesting to note that the volt-ampere product makes no distinction between even and odd junctions as may be seen from the general forms (400), (401), (402), and (403). For the properly terminated structure the relations (406) and (407) give the following volt-ampere ratio

$$\left(\frac{E_0 I_0}{E_k I_k}\right)_{\text{odd or even}} = e^{2k\theta}. \tag{408}$$

These latter results bring out a further point which we shall have occasion to refer to later when discussing filter theory. Namely, if we consider the above structure as consisting of just two networks, we may look upon the combination as one symmetrical network. Its overall propagation function will evidently be 2θ, which, by means of the last three equations, can be expressed as

$$2\theta \doteq \ln\left(\frac{E_0}{E_2}\right) = \ln\left(\frac{I_0}{I_2}\right) = \ln\sqrt{\frac{E_0 I_0}{E_2 I_2}}, \tag{409}$$

for the properly terminated structure. In dealing with symmetrical networks, it often occurs that we wish to employ one-half of such a network, i.e., either half which results from the bisection of the original symmetrical network. If these halves are dissymmetrical, and (409) represents the propagation function for the whole, then for the halves we get the propagation function

$$\left.\begin{aligned} \theta &= \ln\left(\frac{E_0}{E_1}\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}\right) = \ln\left(\frac{E_1}{E_2}\sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}}\right) \\ &= \ln\left(\frac{I_0}{I_1}\sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}}\right) = \ln\left(\frac{I_1}{I_2}\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}\right) \\ &= \ln\sqrt{\frac{E_0 I_0}{E_1 I_1}} \quad = \ln\sqrt{\frac{E_1 I_1}{E_2 I_2}} \end{aligned}\right\} \tag{410}$$

It should, therefore, *not* be inferred that, since the propagation function for each half is half that for the whole network, the logarithmic voltage or current ratios for the halves are half of those for the whole section. By (410) these are

$$\ln\left(\frac{E_0}{E_1}\right) = \theta - \ln\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}};\ \ln\left(\frac{E_1}{E_2}\right) = \theta + \ln\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}, \tag{411}$$

and

$$\ln\left(\frac{I_0}{I_1}\right) = \theta + \ln\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}};\ \ln\left(\frac{I_1}{I_2}\right) = \theta - \ln\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}}. \tag{412}$$

The logarithmic volt-ampere ratios for the halves are half of those for the whole section, however. Where half of a symmetrical network is used, therefore, its effect upon voltage or current alone may not be considered half that of the whole network unless the half is itself symmetrical.

There are many alternate forms in which the image propagation function and characteristic impedances may be expressed. These are obtained by substituting into (387), (392), and (398) from the trans-

formations (295). For the propagation function we may write

$$\left.\begin{aligned}\theta &= \ln(\sqrt{\mathfrak{A}\mathfrak{D}} + \sqrt{\mathfrak{B}\mathfrak{C}}) = \tfrac{1}{2}\ln\left(\frac{\sqrt{y_{11}z_{11}}+1}{\sqrt{y_{11}z_{11}}-1}\right) = \coth^{-1}\sqrt{y_{11}z_{11}}\\ &= \ln\left\{\frac{\sqrt{\sqrt{z_{11}z_{22}}+z_{12}}+\sqrt{\sqrt{z_{11}z_{22}}-z_{12}}}{\sqrt{\sqrt{z_{11}z_{22}}+z_{12}}-\sqrt{\sqrt{z_{11}z_{22}}-z_{12}}}\right\} = \cosh^{-1}\frac{\sqrt{z_{11}z_{22}}}{z_{12}}\end{aligned}\right\}, \tag{413}$$

and for the image impedances

$$Z_{I1} = \sqrt{\frac{z_{11}}{y_{11}}};\ Z_{I2} = \sqrt{\frac{z_{22}}{y_{22}}}. \tag{414}$$

Since $y_{11}z_{11} = y_{22}z_{22}$, the latter product may be substituted in place of $y_{11}z_{11}$ in the forms (413).

It is also interesting to note that the product of the image impedances equals the product of the iterative impedances, namely

$$\sqrt{Z_{I1}Z_{I2}} = \sqrt{Z_{01}Z_{02}} = \sqrt{\frac{\mathfrak{B}}{\mathfrak{C}}} = \sqrt{\frac{z_{11}}{y_{22}}} = \sqrt{\frac{z_{22}}{y_{11}}} \tag{415}$$

When the individual networks in the recurrent structure are symmetrical as evidenced by $\mathfrak{A} = \mathfrak{D}$, then the distinction between the iterative and the image bases ceases to exist, as the reader may readily see by referring to the separate solutions for these cases. We then have

$$\left.\begin{aligned}Z_{I1} = Z_{I2} = Z_{01} = Z_{02} &= \sqrt{\frac{\mathfrak{B}}{\mathfrak{C}}} = \sqrt{\frac{z_{11}}{y_{11}}} = \sqrt{\frac{z_{22}}{y_{22}}}\\ &= (y_{11}^2 - y_{12}^2)^{-1/2} = (z_{11}^2 - z_{12}^2)^{1/2}\end{aligned}\right\}, \tag{416}$$

and

$$\left.\begin{aligned}\theta = \gamma &= \ln(\mathfrak{A} + \sqrt{\mathfrak{A}^2 - 1}) = \ln\left(\frac{(y_{11}-y_{12})^{1/2} + (y_{11}+y_{12})^{1/2}}{(y_{11}-y_{12})^{1/2} - (y_{11}+y_{12})^{1/2}}\right)\\ &= \ln\left(\frac{(z_{11}+z_{12})^{1/2} + (z_{11}-z_{12})^{1/2}}{(z_{11}+z_{12})^{1/2} - (z_{11}-z_{12})^{1/2}}\right) = 2\coth^{-1}\sqrt{\frac{z_{11}+z_{12}}{z_{11}-z_{12}}}\\ &= \cosh^{-1}\left(\frac{z_{11}}{z_{12}}\right)\end{aligned}\right\} \tag{417}$$

There are, of course, an almost endless number of additional variations of these forms which are obtained by applying the transformations (291) to (295) together with the well-known transformations for the hyperbolic and exponential functions.

9. Transmission and attenuation properties of networks. In some of the later chapters it will be our object to discuss the characteristics of special kinds of four-terminal networks known as **filters**. These are net-

works for which the attenuation function (real part of the propagation function) is zero (or nearly zero) for certain frequency ranges, known as the **transmission ranges,** and different from zero (preferably very much different) in the remaining frequency ranges. The latter are referred to as the **attenuation ranges.** The resulting network having such properties is used in communication work wherever it is necessary to select certain frequency ranges as, for example, in connection with carrier or radio systems. Filters are also referred to in general as **selective systems.**

Although it is not our object to discuss this topic in detail here, yet the results which we have derived in the preceding section characteristically indicate what may in general be accomplished in this direction. A knowledge of the fundamental aspects of this situation is helpful in the discussion of certain long-line problems to which we shall come shortly.

Referring to the image propagation function (413), which was determined for the general dissymmetrical case, let us consider more particularly the last logarithmic form given. If, for the moment, we let

$$q = \frac{\sqrt{z_{11}z_{22}} + z_{12}}{\sqrt{z_{11}z_{22}} - z_{12}},$$

then this form may be more briefly written as

$$\theta = \ln\left(\frac{\sqrt{q}+1}{\sqrt{q}-1}\right). \tag{417a}$$

Recognizing that θ will in general be complex, i.e., that

$$\theta = \theta_1 + j\theta_2, \tag{418}$$

we see that θ_1, the attenuation function, will be zero if the magnitude of the argument of the logarithm is unity. This in turn will be the case if q is negative and real, for then the argument will be given by the quotient of a complex number and its negative conjugate. In order for q to be real the z's must be pure reactances. Hence we see that only for a non-dissipative network can the attenuation be zero. Furthermore, the requirement that q be negative leads to the condition

$$-1 \leq \frac{\sqrt{z_{11}z_{22}}}{z_{12}} \leq 1. \tag{419}$$

Thus in a purely reactive structure the attenuation will be zero for those frequency ranges for which the condition (419) is fulfilled. The phase function θ_2 for this region is the angle of

$$\frac{j\sqrt{-q}+1}{j\sqrt{-q}-1},$$

which is given by

$$\theta_2 = -2 \cot^{-1}\sqrt{-q} = \cos^{-1}\left(\frac{q+1}{q-1}\right),$$

or

$$\theta_2 = \cos^{-1}\left(\frac{\sqrt{z_{11}z_{22}}}{z_{12}}\right). \tag{420}$$

This last form may also have been got from (413).

In the frequency ranges for which (419) is not fulfilled, q will be positive and real for the non-dissipative structure. Then $\theta_1 \neq 0$, and since the argument of the logarithm may be either positive or negative, the angle θ_2 will be zero or π, plus or minus an integer number of 2π's. That is, where the attenuation is different from zero, there the phase function remains constant at a value

$$\theta_2 = k\pi, \tag{421}$$

where k may be any positive or negative integer inclusive of zero.

It is of course not strictly correct to call the frequency ranges for which θ_1 is equal to or different from zero the transmission and attenuation regions respectively unless the structure is terminated in its proper image impedance. Unless this is the case, we have seen from the long-line solutions and from the analogous solutions for the uniform recurrent structure that the ratio of input to output depends not alone upon θ but also upon the terminal impedances. However, where the attenuation is high, the terminal conditions have a relatively small effect upon the net behavior. Where the attenuation is small or zero, on the other hand, the question of termination is important in determining the behavior. This will, therefore, become an important item in the discussion of selective systems.

For non-dissipative *symmetrical* networks the condition for zero attenuation and the corresponding phase function are given by

$$\left.\begin{aligned} -1 \leq \frac{z_{11}}{z_{12}} \leq 1 \\ \text{and} \qquad \\ \theta_2 = \cos^{-1}\left(\frac{z_{11}}{z_{12}}\right) \end{aligned}\right\}. \tag{422}$$

10. Propagation and characteristic impedance functions of several common structures. The symmetrical T-, Π-, and lattice-structures occur so frequently in the following work that it will be convenient to have their propagation and characteristic impedance functions determined here.

For the symmetrical T of Fig. 46 we have by (345)

$$z_{11} = \frac{z_1}{2} + z_2; \; z_{12} = z_2, \tag{423}$$

so that

$$z_{11} + z_{12} = \frac{z_1}{2} + 2z_2; \; z_{11} - z_{12} = \frac{z_1}{2}.$$

The application of (416) and (417), therefore, gives

$$Z_T = \sqrt{\frac{z_1}{2}\left(\frac{z_1}{2} + 2z_2\right)} = \sqrt{z_1 z_2}\sqrt{1 + \frac{z_1}{4z_2}}, \tag{424}$$

and

$$\gamma = \ln\left\{\frac{\sqrt{\dfrac{\dfrac{z_1}{2} + 2z_2}{z_1/2}} + 1}{\sqrt{\dfrac{\dfrac{z_1}{2} + 2z_2}{z_1/2}} - 1}\right\} = \cosh^{-1}\left(1 + \frac{z_1}{2z_2}\right). \tag{425}$$

The characteristic impedance is here denoted by Z_T to designate that it refers to the T-structure. The propagation function may be put into an alternative form, which will become more useful later, by noting that

$$\cosh\gamma - 1 = 2\sinh^2\frac{\gamma}{2} = \frac{z_1}{4z_2},$$

so that

$$\gamma = 2\sinh^{-1}\sqrt{\frac{z_1}{4z_2}}. \tag{425a}$$

When the characteristic impedance and propagation functions are given and the component impedances of the corresponding T-structure are desired, these results may be inverted giving

$$\left.\begin{aligned} z_1 &= 2\,Z_T \tanh\frac{\gamma}{2} \\ z_2 &= \frac{Z_T}{\sinh\gamma} \end{aligned}\right\} \tag{426}$$

For the symmetrical Π of Fig. 46 we have by (346)

$$y_{11} = \frac{y_2}{2} + y_1; \; y_{12} = -y_1,$$

from which

$$y_{11} - y_{12} = \frac{y_2}{2} + 2y_1; \; y_{11} + y_{12} = \frac{y_2}{2}.$$

The relations (416) and (417), therefore, give

$$Y_\pi = \sqrt{\frac{y_2}{2}\left(\frac{y_2}{2} + 2y_1\right)} = \sqrt{y_1y_2}\sqrt{1 + \frac{y_2}{4y_1}}, \tag{427}$$

and

$$\gamma = \ln\left\{\frac{\sqrt{\dfrac{\frac{y_2}{2} + 2y_1}{y_2/2}} + 1}{\sqrt{\dfrac{\frac{y_2}{2} + 2y_1}{y_2/2}} - 1}\right\} = \cosh^{-1}\left(1 + \frac{y_2}{4y_1}\right). \tag{428}$$

where Y_π is used to denote the characteristic admittance (reciprocal of the characteristic impedance). The last equation may also be written

$$\gamma = 2 \sinh^{-1}\sqrt{\frac{y_2}{4y_1}}. \tag{428a}$$

Conversely, if γ and Y_π are given, the component admittances of the corresponding Π-structure are found from

$$\left.\begin{aligned} y_1 &= \frac{Y_\pi}{\sinh \gamma} \\ y_2 &= 2Y_\pi \tanh\frac{\gamma}{2} \end{aligned}\right\}. \tag{429}$$

In many practical cases the component admittances in the Π-structure are the reciprocals of the component impedances in the T-structure, i.e., both are derived from the same uniform ladder network. Then

$$y_1 = \frac{1}{z_1};\ y_2 = \frac{1}{z_2},$$

so that

$$Z_\pi = \frac{1}{Y_\pi} = \frac{\sqrt{z_1z_2}}{\sqrt{1 + \frac{z_1}{4z_2}}}. \tag{427a}$$

From a comparison of (425) and (428) or (425*a*) and (428*a*) we see that the propagation functions for the T- and Π-structures for this case are the same.

For the symmetrical lattice of Fig. 47 we substitute into (416) and (417) from (349) and get

$$Z_I = \sqrt{z_az_b} \tag{430}$$

and

$$\gamma = \ln\left(\frac{\sqrt{\frac{z_b}{z_a}} + 1}{\sqrt{\frac{z_b}{z_a}} - 1}\right) = 2 \coth^{-1}\sqrt{\frac{z_b}{z_a}}. \tag{431}$$

The inversion of these gives for the component impedances of the lattice which has Z_I and γ for its characteristic impedance and propagation functions respectively

$$\left.\begin{aligned} z_a &= Z_I \tanh\frac{\gamma}{2} \\ z_b &= Z_I \coth\frac{\gamma}{2} \end{aligned}\right\}. \tag{432}$$

Other structures may be treated in a similar fashion. The above three cases suffice for practically all of our subsequent discussions.

PROBLEMS TO CHAPTER IV

4-1. A bridged T-network according to Fig. 45*a* has the following pure resistance values:

$$z_A = 6;\ z_B = 8;\ z_C = 10;\ 1/y_C = 12.$$

Find the following network equivalents by means of matrix methods:
(*a*) Dissymmetrical T.
(*b*) Dissymmetrical π.
(*c*) Symmetrical T preceded by an ideal transformer.
(*d*) Symmetrical π preceded by an ideal transformer.
(*e*) Symmetrical lattice preceded by an ideal transformer.
(*f*) Symmetrical T followed by an ideal transformer.
(*g*) Symmetrical π followed by an ideal transformer.
(*h*) Symmetrical lattice followed by an ideal transformer.

4-2. A symmetrical four-terminal resistance network has an attenuation of one napier and a characteristic impedance of 1000 ohms.
(*a*) What are the component resistances of a lattice structure having these characteristics?
(*b*) Find the *z*-matrix and from this the equivalent T.
(*c*) Transform this into a ⅂-structure preceded by an ideal transformer, and determine the ratio of the latter.
(*d*) For the structure of part *c* without the ideal transformer, what are the characteristic impedances, and what is the image propagation function?

4-3. For the networks of Problem 4-1 find:
(*a*) The image impedances and image propagation function.
(*b*) The iterative impedances and iterative propagation function.

4-4. A ⅂-structure with series and shunt impedances z_1 and z_2 respectively may theoretically be transformed into an equivalent network consisting of the T-structure of Fig. 44 followed by an ideal transformer having a ratio of $1:a$. If the T-structure thus found is physically realizable (contains positive elements), then by dropping the ideal transformer a change in impedance level at the output end is obtained. (Similarly a change in impedance level can obviously also be obtained at the input end.) Determine the restrictions on the structures of z_1 and z_2 for which this process will be successful, and the maximum range of values for the ratio a which may be obtained in a given case. Find the impedances of the T-structure in terms of z_1, z_2, and a, and determine the final equivalent structure giving the largest change in impedance level.

4–5. If the impedances z_1 and z_2 in the preceding problem are pure inductive reactances, the transformation to an equivalent structure with a change in impedance level at one end may alternatively be carried out by applying the reasoning involved in the discussion on non-dissipative and ideal transformers in section 6. Show that by this line of attack the same results are obtained as by the method of Problem 4–4.

4–6. By expressing the general circuit parameters in terms of the image impedances and the image propagation function thus:

$$\mathcal{A} = \sqrt{Z_{I1}/Z_{I2}} \cosh \theta; \quad \mathcal{B} = \sqrt{Z_{I1}\ Z_{I2}} \sinh \theta;$$

$$\mathcal{C} = \frac{\sinh \theta}{\sqrt{Z_{I1}\ Z_{I2}}}; \qquad \mathcal{D} = \sqrt{Z_{I2}/Z_{I1}} \cosh \theta,$$

show that when a network N is followed in cascade by a network N' and $Z_{I2} = Z_{I1}'$, i.e., the image impedances at the junction are equal, the overall parameters become:

$$\mathcal{A}_0 = \sqrt{Z_{I1}/Z_{I2}'} \cosh (\theta + \theta'); \quad \mathcal{B}_0 = \sqrt{Z_{I1}\ Z_{I2}'} \sinh (\theta + \theta');$$

$$\mathcal{C}_0 = \frac{\sinh (\theta + \theta')}{\sqrt{Z_{I1}\ Z_{I2}'}}; \qquad \mathcal{D}_0 = \sqrt{Z_{I2}'/Z_{I1}} \cosh (\theta + \theta'),$$

from which it follows that under these conditions the overall propagation function is the sum of the individual functions and the image impedances are those of the input to the first and the output of the last network. Show that this result could have been predicted simply on the basis of the definitions of the image impedances and propagation function which follow from the general theory given in the text.

4–7. Show that by starting with the second equation in the y-system:

$$I_2 = y_{21}\ E_1 + y_{22}\ E_2,$$

and noting that $z_{12}/z_{11} = -y_{21}/y_{22}$ and $E_2 = -I_2\ Z_R$ (where Z_R is the terminal impedance), we have in two simple steps:

$$-I_2\left(\frac{1}{y_{22}} + Z_R\right) = \frac{z_{12}}{z_{11}} E_1.$$

The factor of $-I_2$ is the sum of the terminal impedance and that impedance looking back into the output terminals with the input short-circuited; the right-hand side of the equation represents the voltage appearing at the output when that end is open-circuited. A recognition of the relationship thus expressed by the resulting equation is frequently useful in network analysis, and the import of this relationship is known as *Thevenin's theorem*, of which the above constitutes a general proof.

4–8. By choosing for the series and shunt impedances of a ꓶ-structure the inductive and *capacitive* reactances $z_1 = \frac{1}{2} jL_1\omega$ and $z_2 = 1/2jC_2\omega$ respectively, show that the network has a range of free transmission defined by: $0 < \omega < \omega_c$ where $\omega_c = 2/\sqrt{L_1 C_2}$. If we denote $\sqrt{L_1/C_2}$ by R, and ω/ω_c by the variable x, show further that the image impedances become

$$Z_{I1} = \mathrm{R}\sqrt{1 - x^2} \text{ and } Z_{I2} = R/\sqrt{1 - x^2}$$

which are real in the transmission range. The ratio of these impedances is

$$a^2 = Z_{I2}/Z_{I1} = \frac{1}{1 - x^2},$$

which is approximately constant over a sufficiently narrow range of frequencies. This network may, therefore, be used to approximately obtain an impedance transformation without attenuation (ideal transformer characteristics). Show that the variations ΔZ_{I1} and ΔZ_{I2} corresponding to the change Δx are expressed by

$$-\frac{\Delta Z_{I1}}{Z_{I1}} = \frac{\Delta Z_{I2}}{Z_{I2}} = \frac{x}{1 - x^2} \cdot \Delta x$$

and that if we denote the maximum decimal deviation in either impedance from a mean value over the range Δx by ϵ, we have

$$\epsilon = \frac{a^2 - 1}{2} \cdot \frac{\Delta x}{x} = \frac{a^2 - 1}{2} \cdot \frac{\Delta \omega}{\omega},$$

where ω is the mean frequency of the range. Thus show that if we consider a range of frequencies 10 kilocycles per second wide located at the mean frequency of 10^6 cycles per second and demand the impedance ratio $a^2 = 10$, this will be obtained with a maximum tolerance in the impedances of 4.5 per cent. Plot the essential characteristics involved.

4–9. The following data are for a 16-gage telephone cable pair per loop mile and its identical loading coils which are inserted at intervals of 3000 ft:

Frequency cycles/sec	Data per loading coil		Data per loop mile of cable			
	R_L	L_L	R	$L \cdot 10^3$	$C \cdot 10^6$	$G \cdot 10^6$
30	1.10	0.022	42.0	1 060	0.062	0.05
200	1.15	0.022	42.0	1.058	0.062	0.3
500	1.19	0 022	42.1	1.056	0.062	0.7
1,000	1.25	0.022	42.1	1.053	0.062	1.5
2,000	1.60	0.022	42 3	1.046	0.062	3.0
5,000	4 00	0.022	43.2	1.032	0.062	8.8
10,000	12 40	0.022	45.0	1.014	0.062	19.0

Consider a *loading section* to consist of 3000 ft of cable with one-half a loading coil at each end (so-called *mid-coil* section). For this mid-coil loading section determine the equivalent T-network by first finding the equivalent lumped T for the 3000 ft of cable and then adding the loading-coil halves to each series impedance. A 50-mile length of this loaded cable (length between repeater points) consists of 88 of these loading sections in cascade and is terminated at either end in a pure resistance of 800 ohms. For this resultant structure determine the logarithmic insertion ratio for sending-to-receiving-end voltages according to the general relation (130a), p. 75, applied to the cascade of identical symmetrical sections as discussed in section 8. List in separate columns the attenuation loss and phase shift due to the propagation function, the reflection effects, and the interaction effects, and their totals. Plot these three individual contributions as well as the overall attenuation loss and phase shift versus frequency. Plot the real and imaginary components of the characteristic impedance. (Note: The results of this problem are utilized in Chapter XII

for the determination of the requirements of the attenuation and phase corrective networks for this length of loaded cable. See Figs. 221 and 224.)

4–10. Design a pure resistive attenuating network as a single symmetrical lattice structure to have constant image impedances of 800 ohms and attenuation values from 0 to 20 db in 1-db steps. Convert to an equivalent bridge-T design. Alternatively design the attenuator in the form of a number of symmetrical T-sections which may be suitably connected by switches. Determine the best of these designs from the standpoint of economy and practicability.

CHAPTER V

THEOREMS REGARDING DRIVING-POINT IMPEDANCES AND AN EXTENSION TO TWO-TERMINAL PAIRS IN THE REACTIVE CASE

1. **Foster's reactance theorem.** Any linear, passive, two-terminal network is called a **driving-point impedance,** probably on account of the fact that the two terminals are usually specified as that point from which the network is to be driven or excited. The terms *driving-point impedance* and *two-terminal impedance* will hereafter be used indiscriminately. The various impedances which form the branches of the basic four-terminal networks discussed in the previous chapter, such as z_A, z_B, and z_C in the T-structure for example, are driving-point impedances, and may individually be represented by arbitrary linear networks. The further development of four-terminal network theory from this standpoint, therefore, requires that we familiarize ourselves with the characteristics of driving-point impedances in general.

In the design of filters, as well as a number of other special networks, the non-dissipative structures play a predominant part. That is, the theory of operation and design of such four-terminal networks is usually based upon the behavior of structures with negligible ohmic resistances. The presence of the latter in the actual physical network modifies the behavior predicted on the non-dissipative basis only to a small degree. Hence, from the practical standpoint, resistances play only an incidental part in many network design problems. Our discussion of driving-point impedances is, therefore, for the most part restricted to pure reactive networks.

A useful method for determining the physical structure of an arbitrary driving-point reactance from its functional characteristics was generally formulated by Foster.[1] This we wish to discuss now.

Referring to Chapter IV of Volume I, and denoting $j\omega$ by λ for convenience, the mesh and mutual impedances of an arbitrary non-dissipative network of the most general form are given by

$$b_{ik} = L_{ik}\lambda + S_{ik}\lambda^{-1}. \tag{433}$$

If the network is being driven from the first mesh, then the impedance at the generator terminals is given by

$$Z_{11} = \frac{D}{B_{11}}, \tag{434}$$

[1] R. M. Foster, "A Reactance Theorem," B. S. T. J., April, 1924.

in which D is the network determinant and B_{11} the minor of its first row and column. Both D and B_{11} are polynomials in λ. However, due to the fact that negative powers appear in the elements (433), these polynomials will also involve negative powers. The appearance of these is inconvenient in subsequent manipulations. They may be eliminated in the driving-point impedance function (434) by letting

$$b_{ik}^* = \lambda b_{ik} = L_{ik}\lambda^2 + S_{ik}. \tag{435}$$

Then, denoting the determinant $|b_{ik}^*|$ by D^* and its minor by B_{11}^*, we have

$$D^* = \lambda^n D;\ B_{11}^* = \lambda^{n-1}B_{11}, \tag{436}$$

where n equals the number of independent meshes in the network or the number of rows or columns in D. Hence (434) becomes

$$Z_{11} = \frac{D^*}{\lambda B_{11}^*}. \tag{434a}$$

Here D^* and B_{11}^* are polynomials involving only positive powers of λ. In particular, D^* involves powers ranging from $2n$ to zero inclusive, and B_{11}^* involves powers from $2n - 2$ to zero inclusive. Hence the most general driving-point reactance function, representing a network in which every mesh contains independent inductance and capacitance, has the form

$$Z_{11} = \frac{\alpha_{2n}\lambda^{2n} + \alpha_{2n-2}\lambda^{2n-2} + \cdots + \alpha_2\lambda^2 + \alpha_0}{\beta_{2n-1}\lambda^{2n-1} + \beta_{2n-3}\lambda^{2n-3} + \cdots + \beta_3\lambda^3 + \beta_1\lambda}. \tag{437}$$

Note that only even powers of λ appear in the numerator and only odd powers appear in the denominator of this expression.

It is a fundamental law in algebra that polynomials such as these may be factored in terms of their roots. If we consider λ^2 as the variable, then the roots of these polynomials are all negative and real. If we denote them by

$$\lambda^2 = \lambda_k^2 = -\omega_k^2;\ \lambda_k = \pm j\omega_k, \tag{438}$$

and number them so that the indices

$$\left.\begin{aligned} k &= 1, 3, 5, \cdots 2_{n-1} \\ k &= 0, 2, 4, \cdots 2_{n-2} \end{aligned}\right\} \tag{439}$$

refer to the roots of the numerator and denominator respectively, where $\lambda_0 = 0$ represents the zero root of the denominator, then (437) may be written

$$Z_{11} = H \cdot \frac{(\lambda^2 - \lambda_1^2)(\lambda^2 - \lambda_3^2) \cdots (\lambda^2 - \lambda^2_{2n-1})}{\lambda(\lambda^2 - \lambda_2^2)(\lambda^2 - \lambda_4^2) \cdots (\lambda^2 - \lambda^2_{2n-2})}. \tag{440}$$

Here H is the quotient of the first coefficients in the two polynomials, i.e.

$$H = \frac{\alpha_{2n}}{\beta_{2n-1}}. \tag{441}$$

Substituting $j\omega$ for λ again (440) becomes

$$Z_{11} = j\omega H \cdot \frac{(\omega^2 - \omega_1^2)(\omega^2 - \omega_3^2) \cdots (\omega^2 - \omega^2{}_{2n-1})}{\omega^2(\omega^2 - \omega_2^2)(\omega^2 - \omega_4^2) \cdots (\omega^2 - \omega^2{}_{2n-2})}. \tag{440a}$$

This factored form places the poles and zeros of the reactance function in evidence, and emphasizes the fact that they are *simple.*[1] Since the numerator of (437) or (440*a*) is one degree higher in λ or ω than the denominator, we see that Z_{11} has a pole at infinity in addition to having one at the origin. This is so in the general case where each mesh contains an independent inductance and capacitance. Special cases where the origin or infinity may be zeros will be discussed later.

This result shows that *a driving-point reactance function is uniquely specified by the location of its internal*[2] *zeros and poles plus one additional piece of information which in (440a) takes the form of a multiplying factor H.* Thus, if we are given the frequencies ω_1, ω_3, $\cdots$ at which a reactance is to be zero, and the frequencies ω_2, ω_4, $\cdots$ at which it is to be infinite, together with the value which it is to have at one additional frequency (this is equivalent to having H specified), then we can write down the analytic expression in the form (440*a*), or its alternatives (437) or (440), which may represent a physical network having these characteristics.

The driving-point reactance function for a physical network, therefore, has a rather restricted form. We shall see later that it possesses another important property, namely that its slope versus frequency is everywhere positive,[3] i.e.,

$$\frac{dZ_{11}}{jd\omega} > 0, \text{ for } -\infty < \omega < \infty. \tag{442}$$

At a pole the function must, therefore, change sign, as may also be seen by inspection of (440*a*). *The positiveness of the slope of Z_{11} requires that*

[1] A proof of this is given in the following chapter. See footnote 1 on p. 229.

[2] Poles or zeros at the origin or at infinity are referred to as *external*, and play no part in the specification of a reactance function, as will become clear from the following discussion.

[3] A general proof based upon energy considerations is given in section 3 of the following chapter.

the poles and zeros alternate, because two successive poles or zeros would require that the slope be negative over part of the intervening frequency region. The alternation of poles and zeros is expressed by

$$0 = \omega_0 < \omega_1 < \omega_2 < \cdots < \omega_{2n-2} < \omega_{2n-1} < \infty, \qquad (443)$$

and is called the **separation property**[1] of the zeros and poles of a physically realizable reactance function. Fig. 55 illustrates the general appearance of such a function.

The problem with which Foster deals is that of finding a network representation for the reactance function. Here either the function itself may be given in the form (440*a*), let us say, or the data necessary for the analytic formulation of this function may be given. Note that

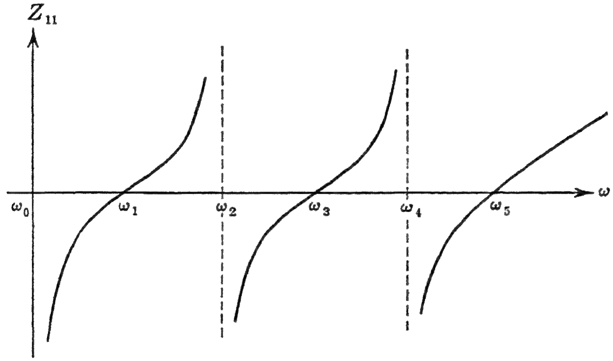


FIG. 55.—General appearance of the driving-point reactance function of a physical network.

the given data cannot be arbitrary. The prescribed zeros and poles must separate each other; otherwise no corresponding physical network exists. Furthermore, the prescribed value at a given frequency must have the proper sign so that (442) will not be violated. That is, if the prescribed value lies immediately to the right of a pole it must be negative, otherwise positive, as may be appreciated by referring to Fig. 55. This is equivalent to saying that H must be positive.

In practice the given data may not always be in this form. The desired reactance function may be required to follow a given curve within specified frequency limits, and no location of zeros and poles outside this region may be given. This case is much more difficult to

[1] An independent proof of this property is to be found in "The Advanced Part of a Treatise on the Dynamics of a System of Rigid Bodies," by Edward John Routh, Macmillan & Co., 1884, pp. 36–38.

treat since the number and location of zeros and poles must first be found such that the given curve in the specified frequency region is approximated with a sufficient degree of accuracy. The extent of the resulting network will then depend upon the degree of approximation required. Since the general solution to this problem is at present still in the cut-and-try state, we shall not take up its discussion here, but rather continue with the simpler problem formulated above.

The form (437) or (440a) may be identified with a physical network in several ways. If we attempt a partial fraction expansion of this function, with the hope that we will thus represent the desired network by a series combination of simpler ones, the expression (440a) will take the form

$$Z_{11} = j\omega H\left\{1 + \frac{A_0}{\omega^2} + \frac{A_2}{\omega^2 - \omega_2^2} + \cdots + \frac{A_{2n-2}}{\omega^2 - \omega^2_{2n-2}}\right\}. \qquad (444)$$

The term 1 is necessary in order that this function shall approach the value $j\omega H$ for $\omega \rightarrow \infty$ as required by inspection of (440a); i.e., this term must account for the pole at infinity just as the succeeding terms individually account for the poles at $\omega = 0, \omega_2, \omega_4, \cdots$.

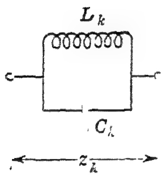


FIG. 56.—Antiresonant component whose reactance function may analytically represent a term in the partial fraction expansion of eq. (444).

The coefficients $A_0 \cdots A_{2n-2}$ may be evaluated by equating the expansion (444) to the original form (440a). For the determination of any coefficient A_k, consider a frequency very close to ω_k. For this frequency the term with A_k becomes extremely large so that, as ω approaches ω_k, all other terms in (444) become negligible in comparison with the A_k term. In the limit $\omega \rightarrow \omega_k$ this will be rigorously correct. At the same time the expression (440a) also becomes large on account of the factor $(\omega^2 - \omega_k^2)$ in its denominator. Equating (440a) to (444), which is represented by its A_k term alone, the factors $(\omega^2 - \omega_k^2)$ may be cancelled so that the limit gives

$$A_k = \frac{(\omega_k^2 - \omega_1^2)(\omega_k^2 - \omega_3^2) \cdots (\omega_k^2 - \omega^2_{2n-1})}{\omega_k^2(\omega_k^2 - \omega_2^2) \cdots (\omega_k^2 - \omega^2_{k-2})(\omega_k^2 - \omega^2_{k+2}) \cdots (\omega_k^2 - \omega^2_{2n-2})}. \qquad (445)$$

The form in which the denominator is written is intended to show that the factor $(\omega_k^2 - \omega_k^2)$ is absent.

Returning to the expansion (444) we note that the first two terms have the form of the reactances of an inductance and a capacitance respectively. For the identification of the rest, let us consider the net-

work of Fig. 56 consisting of an inductance and capacitance in parallel. The impedance of this component network is

$$z_k = \frac{jL_k\omega}{1 - L_kC_k\omega^2} = \frac{j\omega\left(-\dfrac{1}{C_k}\right)}{\omega^2 - \omega_k^2}, \tag{446}$$

where

$$\omega_k^2 = \frac{1}{L_kC_k} \tag{447}$$

is denoted as the **anti-resonant** frequency. The impedance z_k itself is spoken of as an **anti-resonant component** since at the frequency ω_k it becomes infinite. Such a component as this represents a pole at $\omega = \omega_k$

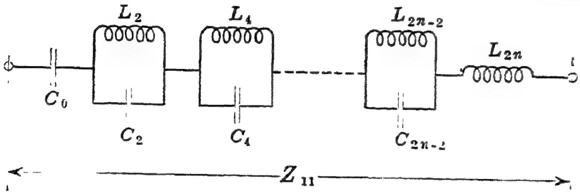


FIG. 57.—Foster form for a driving-point reactance obtained by a partial fraction expansion of the reactance function.

and hence should also represent any of the remaining terms in (444). This will evidently be so if we make

$$-\frac{1}{C_k} = H \cdot A_k. \tag{448}$$

Substitution of (445) into (448), therefore, determines the component networks.[1] In order to avoid writing the cumbersome form (445), we note that this is the same as Z_{11} with the omission of $j\omega H$ and the factor $(\omega^2 - \omega_k^2)$, evaluated at the frequency ω_k. Thus we obtain for (448)

$$C_k = -\left[\frac{j\omega Y_{11}}{\omega^2 - \omega_k^2}\right]_{\omega=\omega_k}; \ (k = 0, 2, 4, \cdots 2n - 2), \tag{449}$$

where the admittance function

$$Y_{11} = Z_{11}^{-1} \tag{450}$$

has been used in order to make the expression more compact.

With (447) and (449) the final network representation for the given function is that shown in Fig. 57 where

$$L_{2n} = H \tag{451}$$

[1] The reader should note that the A_k's according to (445) will be negative provided the poles and zeros separate each other. Thus (448) shows that the C_k's, and hence the L_k's will be positive (i.e., the network will be realizable in the form of anti-resonant components in series) provided H is positive and the separation property is fulfilled.

accounts for the first term in (444). Thus we have shown that an arbitrary two-terminal reactive network is always reducible to this form, which consists of anti-resonant components in series. In this sense the condenser C_0 and the coil L_{2n} should be thought of as special cases of anti-resonant components for which the anti-resonant frequencies are zero and infinity respectively.

The analytic expression (446) may, therefore, be considered as representing any component. Differentiating this with respect to ω we find

$$\frac{dz_k}{jd\omega} = \frac{L_k(1 + L_kC_k\omega^2)}{(1 - L_kC_k\omega^2)^2} \tag{452}$$

which is evidently positive for all frequencies. Since the sum of such derivatives must, of course, also be positive, the property expressed by (442) is thus shown to hold for the network form of Fig. 57.

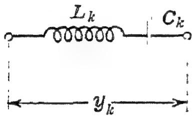


FIG. 58.—Resonant component whose reactance function may analytically represent a term in the partial fraction expansion of the driving point susceptance function.

This form for the physical realization of the reactance function (440a) is by no means the only one. Another is found by treating the admittance function (450) in a similar manner, i.e., expanding it into partial fractions and identifying the terms in this with simple component networks. This expansion, which has the form

$$Y_{11} = -j\omega H^{-1}\left\{\frac{B_1}{\omega^2 - \omega_1^2} + \frac{B_3}{\omega^2 - \omega_3^2} + \cdots + \frac{B_{2n-1}}{\omega^2 - \omega^2_{2n-1}}\right\}, \tag{453}$$

places the poles of Y_{11} or the zeros of Z_{11} in evidence.

The coefficients may be evaluated in the same manner as in the preceding case, giving

$$B_k = \left[\frac{H(\omega^2 - \omega_k^2)}{-j\omega Z_{11}}\right]_{\omega=\omega_k}; \ (k = 1, 3, \cdots 2n - 1). \tag{454}$$

Each term in (453) may now be identified with a component network. For the one shown in Fig. 58 we have

$$y_k = \frac{jC_k\omega}{1 - L_kC_k\omega^2} = \frac{j\omega\left(-\frac{1}{L_k}\right)}{\omega^2 - \omega_k^2}, \tag{455}$$

where

$$\omega_k^2 = \frac{1}{L_kC_k} \tag{456}$$

is the **resonant frequency.** This is the frequency at which the impedance of this network vanishes. The network is, therefore, referred to as a **resonant component.**

Comparing (455) with the terms in (453) we recognize that

$$\frac{1}{L_k} = H^{-1} \cdot B_k, \tag{457}$$

so that with (454) we have

$$L_k = -\left[\frac{j\omega Z_{11}}{\omega^2 - \omega_k^2}\right]_{\omega=\omega_k}; \ (k = 1, 3, \cdots 2n - 1). \tag{458}$$

The network representing Y_{11} or Z_{11} is, therefore, that shown in Fig. 59.

2. Function theoretical point of view. Foster's reactance theorem takes on a more general significance when interpreted in the light of the theory of functions of a complex variable. The driving-point impedance $Z_{11}(\lambda)$ may be considered in general as a function of the complex variable $\lambda = \gamma + j\omega$,[1] although the study of its behavior as a function of frequency involves pure imaginary values of λ only. In the reactive case the poles and zeros of $Z_{11}(\lambda)$ lie along the imaginary axis and hence they are realized by real frequencies. In the dissipative case the poles and zeros do not lie along the imaginary axis but are confined to the *left half-plane*, i.e., to that region of the complex plane which lies to the left of the imaginary axis. That this is so may be seen if we recall that $Z_{11}(\lambda) = 0$ is the determinantal equation of the network. Since the roots of this equation must have negative real parts, it is clear that the zeros of $Z_{11}(\lambda)$ lie in the left half-plane. Similarly $Y_{11}(\lambda) = 0$ is the determinantal equation of the abbreviated network which results by dropping the first mesh, since this is the same as $B_{11}(\lambda) = 0$, the latter being the minor of the network determinant for the first row and column. Thus the poles of $Z_{11}(\lambda)$ also lie in the left half-plane. For the non-dissipative case, however, the roots of these determinantal equations become pure imaginary quantities, i.e., all decrements are zero, and hence the zeros and poles lie along the imaginary axis. This is the case to which Foster's theorem applies.

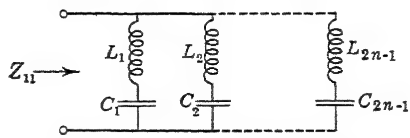


FIG. 59.—Alternative Foster form for realizing a driving-point reactance as a parallel combination of the resonant components shown in Fig. 58.

It is a well-known fact in function theory that a function may be expanded in a *Laurent series* about any one of its singularities. In particular, if the singularity is a pole of finite multiplicity, then the negative branch of the expansion will be restricted to a finite number of terms.

[1] The symbol γ is here used to denote the real part of λ. It is not likely to be confused with the propagation function which does not occur in the present chapter.

Thus for the function $f(\lambda)$ about a pole at $\lambda = \lambda_1$ the Laurent expansion is

$$f(\lambda) = \sum_{k=0}^{\infty} a_k(\lambda - \lambda_1)^k + \frac{a_{-1}}{\lambda - \lambda_1} + \cdots + \frac{a_{-s}}{(\lambda - \lambda_1)^s}, \tag{459}$$

where s is the *multiplicity* of the pole. The portion with descending powers is the **principal part** and represents the effect of the pole. Denoting this part by $h_1(\lambda)$ we have

$$f(\lambda) - h_1(\lambda) = \sum_{k=0}^{\infty} a_k(\lambda - \lambda_1)^k, \tag{459a}$$

which is the ordinary *Taylor expansion* of the function about $\lambda = \lambda_1$ after the pole at this point has been removed. Its radius of convergence extends to the next nearest pole.

If $f(\lambda)$ is *rational,* i.e., has a finite number of poles, then a Laurent expansion can be written at each of these. If all the principal parts of these expansions are subtracted from the original function, the remainder will be regular in the entire λ-plane with the possible exception of the point ∞. Denoting this remainder by $g(\lambda)$, and the principal parts at the points $\lambda = \lambda_1, \lambda_2, \cdots \lambda_n$ by $h_1, h_2, \cdots h_n$, we have

$$f(\lambda) = h_1(\lambda) + h_2(\lambda) + \cdots + h_n(\lambda) + g(\lambda), \tag{460}$$

which is regarded as the **partial fraction expansion** of $f(\lambda)$. If the latter has a pole at infinity also, then $g(\lambda)$ contains this pole.

Regarding our reactance function $Z_{11}(\lambda)$, we have seen that its poles are all *simple.* This means that its principal parts consist of the first negative terms only, and that $g(\lambda)$ likewise consists of the first positive term only since the pole at infinity, if present, must be simple, and no zero order term can exist on account of the pure imaginary character of $Z_{11}(\lambda)$. Thus the partial fraction expansion of the reactance function must have the form

$$Z_{11}(\lambda) = \frac{k_0}{\lambda} + \frac{k_2}{\lambda - \lambda_2} + \frac{k_2'}{\lambda + \lambda_2} + \cdots + \frac{k_{2n-2}}{\lambda - \lambda_{2n-2}} + \frac{k'_{2n-2}}{\lambda + \lambda_{2n-2}} + k_{2n}\lambda, \tag{461}$$

where, with the exception of the first and last, the terms are in the form of conjugate pairs. That is, since $\lambda = \pm \lambda_k$ are a conjugate pair of poles, k_k and k_k' are a conjugate pair of coefficients.

The first and last terms in (461) represent the poles at the origin and at infinity respectively. Their coefficients may be evaluated from (437) directly. For very small values of λ the terms of lowest degree in the two polynomials alone represent the function, while for very large

values of λ the terms of highest degree only need be retained. Thus it is clear that

$$k_0 = \frac{\alpha_0}{\beta_1};\ k_{2n} = \frac{\alpha_{2n}}{\beta_{2n-1}}. \tag{462}$$

The remaining k's are the coefficients of the first negative terms in the Laurent expansions of $Z_{11}(\lambda)$ about the respective poles. These may be most easily obtained by expanding the admittance Y_{11} in a Taylor's series about the same points, which may be done because these are zeros for Y_{11}, and then getting the desired Laurent expansion by dividing the Taylor's series into unity. For the Taylor expansion we have

$$Y_{11}(\lambda) = Y_{11}(\lambda_k) + \frac{1}{1!}\left(\frac{dY_{11}}{d\lambda}\right)_{\lambda=\lambda_k} \cdot (\lambda - \lambda_k) + \frac{1}{2!}\left(\frac{d^2Y_{11}}{d\lambda^2}\right)_{\lambda=\lambda_k} \cdot (\lambda - \lambda_k)^2 \cdots \tag{463}$$

Since $Y_{11}(\lambda_k) = 0$, the reciprocal of this series will start with the term

$$\left(\frac{dY_{11}}{d\lambda}\right)^{-1}_{\lambda=\lambda_k} \cdot (\lambda - \lambda_k)^{-1}. \tag{464}$$

The coefficient of this term is the desired k_k, i.e.,

$$k_k = \left(\frac{dY_{11}}{d\lambda}\right)^{-1}_{\lambda=\lambda_k};\ (k = 0, 2, 4, \cdots 2n - 2). \tag{465}$$

These k's are called the **residues** of $Z_{11}(\lambda)$ in the respective poles.

Because $Y_{11}(\lambda) = Y_{11}(j\omega)$ is a pure imaginary in the reactive case, and

$$\frac{dY_{11}}{d\lambda} = \frac{dY_{11}}{jd\omega},$$

it follows that k_k is a *real* number. On account of the positive slope of $Z_{11}(\lambda)$ and $Y_{11}(\lambda)$, it is also *positive*.[1] This means that the conjugate value k_k' is equal to k_k, i.e.,

$$k_k' = k_k. \tag{466}$$

Making use of this in (461) by combining corresponding pairs of terms we find

$$Z_{11}(\lambda) = \frac{k_0}{\lambda} + \frac{2k_2\lambda}{\lambda^2 - \lambda_2^2} + \frac{2k_4\lambda}{\lambda^2 - \lambda_4^2} + \cdots + \frac{2k_{2n-2}\lambda}{\lambda^2 - \lambda^2_{2n-2}} + k_{2n}\lambda. \tag{461a}$$

[1] The positive and real character of these residues is thus shown to be equivalent to the separation property pointed out above. The condition for physical realizability may, therefore, be stated by saying that *the poles of Z_{11} must be simple and the residues of Z_{11} in its poles must be real and positive, i.e.,* $k_k \geq 0$. That this is sufficient is evident from the relations (467) and (468).

The first and last terms obviously represent the capacitance and inductance

$$\left.\begin{aligned} C_0 &= \frac{1}{k_0} \\ \text{and} \qquad L_{2n} &= k_{2n} \end{aligned}\right\}, \tag{467}$$

in the network of Fig. 57, while the remaining terms represent anti-resonant components for which by (446)

$$C_k = \frac{1}{2k_k}; \ (k = 2, 4, \cdots 2n - 2). \tag{468}$$

If we use the reciprocal of (440) for $Y_{11}(\lambda)$ and form the derivative at any of its zeros we find

$$\left.\begin{aligned} \left(\frac{dY_{11}(\lambda)}{d\lambda}\right)_{\lambda=\lambda_k} = \left(\frac{dY_{11}(j\omega)}{jd\omega}\right)_{\omega=\omega_k} = -\left[\frac{2j\omega Y_{11}}{\omega^2 - \omega_k^2}\right]_{\omega=\omega_k} \\ \text{for} \qquad (k = 2, 4, \cdots 2n - 2), \end{aligned}\right\} \tag{469}$$

and

$$\left.\begin{aligned} \left(\frac{dY_{11}(\lambda)}{d\lambda}\right)_{\lambda=\lambda_k} = -\left[\frac{j\omega Y_{11}}{\omega^2 - \omega_k^2}\right]_{\omega=\omega_k} \\ \text{for} \qquad k = 0, \text{ i.e., } \omega = \omega_0 = 0. \end{aligned}\right\} \tag{469a}$$

Thus the results (467) and (468) are found to agree with the previous results (449) and (451) for this case.

Similarly the admittance function for the general case when expanded into partial fractions gives

$$Y_{11}(\lambda) = \frac{k_1}{\lambda - \lambda_1} + \frac{k'_1}{\lambda + \lambda_1} + \cdots + \frac{k_{2n-1}}{\lambda - \lambda_{2n-1}} + \frac{k'_{2n-1}}{\lambda + \lambda_{2n-1}}, \tag{470}$$

where

$$k_k = k_k' = \left(\frac{dZ_{11}}{d\lambda}\right)^{-1}_{\lambda=\lambda_k}; \ (k = 1, 3, \cdots 2n - 1) \tag{471}$$

are the residues of $Y_{11}(\lambda)$ in its poles $\lambda_1 \cdots \lambda_{2n-1}$, so that (470) becomes

$$Y_{11}(\lambda) = \frac{2k_1\lambda}{\lambda^2 - \lambda_1^2} + \frac{2k_3\lambda}{\lambda^2 - \lambda_3^2} + \cdots + \frac{2k_{2n-1}\lambda}{\lambda^2 - \lambda^2_{2n-1}}. \tag{470a}$$

By means of (455) this is identified with the network of Fig. 59 where

$$L_k = \frac{1}{2k_k}; \ (k = 1, 3, \cdots 2n - 1). \tag{472}$$

Noting that by (440) we get

$$\left.\begin{aligned} \left(\frac{dZ_{11}(\lambda)}{d\lambda}\right)_{\lambda=\lambda_k} = \left(\frac{dZ_{11}(j\omega)}{jd\omega}\right)_{\omega=\omega_k} = -\left[\frac{2j\omega Z_{11}}{\omega^2 - \omega_k^2}\right]_{\omega=\omega_k} \\ \text{for} \qquad (k = 1, 3, 5, \cdots 2n - 1), \end{aligned}\right\} \tag{473}$$

we see that the present result agrees with (458).[1]

[1] The condition for physical realizability is again stated by $k_k \geq 0$ and real.

3. Special cases. So far we have spoken of the general case where each mesh in the network independently contains inductance and capacitance. This leads to a driving-point reactance function which possesses poles at the origin and at infinity as is evident from (461), for example, on account of the presence of the terms with k_0 and k_{2n} respectively. Special cases occur for which either or both of these coefficients are zero. According to (462) these cases will lead to polynomial forms (437) for $Z_{11}(\lambda)$ in which either or both α_0 and α_{2n} are zero. Note, however, from (461a) that when $k_0 = 0$, Z_{11} has a zero at the origin, and when $k_{2n} = 0$, it has a zero at infinity. In these cases Y_{11} has a pole at the origin or at infinity.

The treatment of these special cases is the same as for the general one. For the network representation of Fig. 57, when k_0 or k_{2n} are zero, then by (467) $C_0 = \infty$ or $L_{2n} = 0$, respectively, i.e., these elements which account for the poles at zero and at infinity are absent in the resulting network. The relation (449) still holds for the remaining antiresonant components.

For the network representation of Fig. 59, when $k_0 = 0$ ($\alpha_0 = 0$), then $Y_{11}(\lambda)$ has a pole at the origin. Its partial fraction expansion, instead of being given by (470a), is

$$Y_{11}(\lambda) = \frac{k_1}{\lambda} + \frac{2k_3\lambda}{\lambda^2 - \lambda_3^2} + \cdots + \frac{2k_{2n-1}\lambda}{\lambda^2 - \lambda^2_{2n-1}}, \tag{470b}$$

where the k's are still given by (471). The inductances are, therefore, given by (458) with the reminder that $\omega_1 = 0$ and hence by (456) that $C_1 = \infty$, i.e., this capacitance is absent in the resulting network.

When, for this same network representation, we have $k_{2n} = 0$ ($\alpha_{2n} = 0$), then $Y_{11}(\lambda)$ has a pole at infinity. Its partial fraction expansion then has the form

$$Y_{11}(\lambda) = \frac{2k_1\lambda}{\lambda^2 - \lambda_1^2} + \cdots + \frac{2k_{2n-3}\lambda}{\lambda^2 - \lambda^2_{2n-3}} + k_{2n-1}\lambda, \tag{470c}$$

where, with the exception of k_{2n-1}, the k's are still given by (471). By referring to (437) and considering its reciprocal for $\alpha_{2n} = 0$, it becomes clear that

$$k_{2n-1} = \frac{\beta_{2n-1}}{\alpha_{2n-2}}. \tag{474}$$

The term in (470c) with this coefficient represents a capacitance alone with the value $C_{2n-1} = k_{2n-1}$. The corresponding inductance L_{2n-1} is zero since $\omega_{2n-1} \to \infty$. The remaining inductances are given by (458).

4. Illustrations. Foster's reactance theorem not only gives us a means for physically realizing a driving-point reactance function from pre-

scribed characteristics, but also abbreviates the process of determining the analytic form of the reactance function for a given non-dissipative two-terminal network. Consider, for example, the network of Fig. 60. This is a four-mesh network. Using the method outlined in Section 7, Chapter V, of Volume I, for the determination of the degree of the determinantal equation, we see that $Z_{11} = 0$ has seven roots. Six of these are in the form of conjugate imaginary pairs and hence represent three zeros for Z_{11} between zero and infinity. These are referred to as **internal zeros** to distinguish them from **external zeros** which may occur at *zero* or *infinity*. The seventh root of $Z_{11} = 0$ represents such an external zero. By inspection of Fig. 60 we see that this is located at the origin.[1]

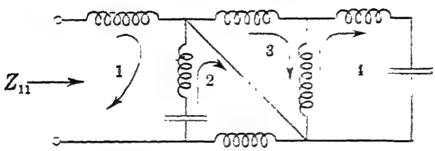


FIG. 60.—Example of an arbitrary driving-point reactance.

Hence the distribution of zeros and poles for this reactance function is as shown in Fig. 61 which represents a frequency axis on which the locations of zeros and poles are indicated by circles and crosses respectively, and infinity is represented as a finite point. In writing the analytic expression for Z_{11} we note from the general form (440*a*) that internal zeros and poles are indicated by factors $(\omega^2 - \omega_k^2)$ in the numerator or denominator respectively, while external zeros or poles involve merely the factor $j\omega$. The sign is determined so that the slope will be positive. Thus we get in this case

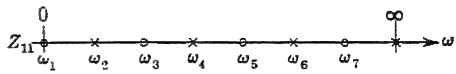


FIG. 61.—Distribution of zeros and poles for the reactance function represented by the network of Fig. 60.

$$Z_{11} = j\omega H \cdot \frac{(\omega^2 - \omega_3^2)(\omega^2 - \omega_5^2)(\omega^2 - \omega_7^2)}{(\omega^2 - \omega_2^2)(\omega^2 - \omega_4^2)(\omega^2 - \omega_6^2)}. \tag{475}$$

The reader should note that this form behaves properly both at the origin and at infinity.

This method of setting up the reactance function is applicable to any case and is much shorter than the ordinary network method. The

[1] When the degree of the determinantal equation is odd, one root is either zero or infinite. Inspection of the network readily shows whether the corresponding zero of Z_{11} lies at the origin or at infinity. When the degree is even the zeros of Z_{11} are either all internal (the number being equal to half the degree of the determinantal equation) or two of them are external (one at the origin and one at infinity). The external zeros are simple roots while the internal ones are given by pairs of conjugate imaginaries.

locations of the zeros and poles must, of course, be determined from the equations

$$\left.\begin{aligned} D &= 0 \\ \text{and} \qquad B_{11} &= 0 \end{aligned}\right\}, \tag{476}$$

where D is the network determinant and B_{11} its minor for the first row and column. In a numerical example, this involves the usual difficulties encountered in the solution of such equations.

Suppose now that the zeros and poles are prescribed, namely

$$\left.\begin{aligned} \omega_3 &= 300, & \omega_2 &= 200 \\ \omega_5 &= 500, & \omega_4 &= 400 \\ \omega_7 &= 700, & \omega_6 &= 600 \end{aligned}\right\}. \tag{477}$$

In addition, the reactance shall equal, say, $j1250$ at $\omega = 50$. In order to get the physical form shown in Fig. 57 we note that the three internal poles require three anti-resonant components, while the pole at infinity requires a coil. The condenser C_0 is absent because the origin is a zero. Thus we get the network of Fig. 62.

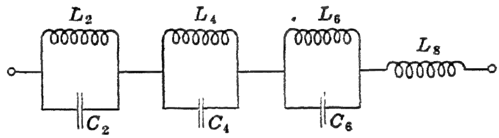


FIG. 62.—Foster representation of the driving-point reactance of Fig. 60 as a series combination of anti-resonant components.

Before proceeding with the evaluation of the network parameters, we must determine H. Substituting $\omega = 50$ into (475) and equating to $j1250$ gives

$$j1250 = j50\,H \cdot \frac{(0.25 - 9)(0.25 - 25)(0.25 - 49)}{(0.25 - 4)(0.25 - 16)(0.25 - 36)},$$

from which

$$H = 5 = L_8.$$

Using (449) for $k = 2, 4, 6$, we get

$$C_2 = -\frac{1}{5}\,\frac{(4 - 16)(4 - 36)\cdot 10^4}{(4 - 9)(4 - 25)(4 - 49)} = 1.625\cdot 10^{-6},$$

$$C_4 = -\frac{1}{5}\,\frac{(16 - 4)(16 - 36)\cdot 10^4}{(16 - 9)(16 - 25)(16 - 49)} = 2.310\cdot 10^{-6}.$$

$$C_6 = -\frac{1}{5}\,\frac{(36 - 4)(36 - 16)\cdot 10^4}{(36 - 9)(36 - 25)(36 - 49)} = 3.315\cdot 10^{-6}.$$

For the corresponding inductances we have by (447)

$$L_2 = 15.38;\; L_4 = 2.704;\; L_6 = 0.838.$$

For the evaluation of the physical form shown in Fig. 59 we note that the three internal zeros call for three resonant components, and the zero at the origin requires a coil. Hence the network is that shown in Fig. 63. The parameter values are found from (458) using (475) and $H = 5$. We thus have

$$L_1 = 5\frac{9 \cdot 25 \cdot 49}{4 \cdot 16 \cdot 36} = 23.92,$$

$$L_3 = 5\frac{9(9-25)(9-49)}{(9-4)(9-16)(9-36)} = 30.5,$$

$$L_5 = 5\frac{25(25-9)(25-49)}{(25-4)(25-16)(25-36)} = 23.1,$$

$$L_7 = 5\frac{49(49-9)(49-25)}{(49-4)(49-16)(49-36)} = 12.18.$$

The capacitances corresponding to the last three are

$$C_3 = 0.364 \cdot 10^{-6};\ C_5 = 0.173 \cdot 10^{-6};\ C_7 = 0.167 \cdot 10^{-6}.$$

5. Least number of elements. An additional point of interest is brought out by the example just discussed. This concerns itself with the number of network elements involved in the physical networks representing a driving-point reactance function. Here we see that the network of Fig. 60 involves 9 elements whereas the equivalent networks of Figs. 62 and 63 involve only 7. The networks which are obtained by Foster's method always contain the *least* number of elements. They are, therefore, called the **fundamental** or **canonic forms.** The network of Fig. 60 apparently contains *two* superfluous elements.

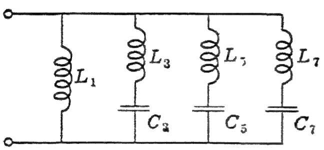


FIG. 63.—Alternative Foster representation of the driving-point reactance of Fig. 60 as a parallel combination of resonant components.

The results of Foster's reactance theorem show that *the least number of elements by which a given driving-point reactance function may be realized equals one more than the sum of its internal poles and zeros.* This fact will become important in the development of filter theory to be discussed later.

6. Cauer's extension of Foster's theorem. The reactance function (437) may be realized physically in two additional fundamental forms. These were first pointed out by Cauer.[1] Their method of realization depends upon a continued fraction expansion of (437).

[1] W. Cauer, "Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit," Archiv f. Elektrotechnik, Vol. 17, p. 355, 1927.

It is well known that the impedance of a ladder structure as illustrated in Fig. 64 is expressible in a continued fraction. This may easily be seen by forming the expression for the impedance Z by alternate series and parallel combinations of its components, starting with the right-hand end and working back toward the terminals. With the series arms expressed as impedances and the shunt arms as admittances, this takes the form

$$Z = z_1 + \cfrac{1}{y_2 + \cfrac{1}{z_3 + \cfrac{1}{y_4 + \cdots + \cfrac{1}{z_{n-1} + \cfrac{1}{y_n}}}}}. \tag{478}$$

The quotient of polynomials (437) may be expanded in such a continued fraction in two general ways. One of these begins by dividing

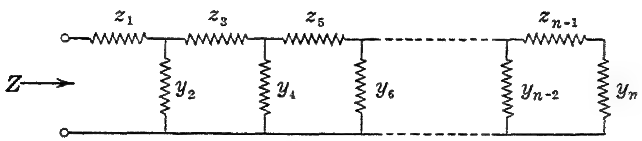


FIG. 64.—Ladder structure whose driving-point impedance is expressible as a continued fraction according to eq. 478.

the denominator into the numerator as if by long division but breaking off after the first term of the quotient is obtained, then inverting the remainder and dividing out one more term, inverting again and dividing again, and so on until the process terminates of itself. If we let

$$\left. \begin{aligned} P &= \alpha_{2n}\lambda^{2n} + \cdots + \alpha_0 \\ Q &= \beta_{2n-1}\lambda^{2n-1} + \cdots + \beta_1\lambda \end{aligned} \right\}, \tag{479}$$

then

$$Z_{11} = \frac{P}{Q}.$$

Beginning the division with the highest terms in P and Q we have

$$Z_{11} = a_1\lambda + \frac{P'}{Q},$$

where P'/Q is the remainder. P' is a polynomial of the $(2n-2)$th

degree. Hence Q is one degree higher than P'. Inverting the remainder and dividing out one more term, therefore, gives

$$Z_{11} = a_1\lambda + \cfrac{1}{b_1\lambda + \cfrac{Q'}{P'}},$$

where now Q' is of the $(2n-3)$th degree, i.e., one degree lower than P'. Each division and inversion, therefore, reduces the remaining quotient by one degree. Indicating the degrees of the polynomials only, the original quotient and subsequent inverted remainders are

$$\frac{2n}{2n-1};\ \frac{2n-1}{2n-2};\ \frac{2n-2}{2n-3};\ \cdots\ \frac{3}{2};\ \frac{2}{1};\ \frac{1}{0}.$$

Thus it is clear that the process will terminate after $2n$ terms. The continued fraction expansion of (437), therefore, reads

$$Z_{11} = a_1\lambda + \cfrac{1}{b_1\lambda + \cfrac{1}{a_2\lambda + \cfrac{1}{b_2\lambda + \cdots + \cfrac{1}{a_n\lambda + \cfrac{1}{b_n\lambda.}}}}} \tag{480}$$

Comparing this with (478) we see that

$$a_k\lambda = z_k;\ b_k\lambda = y_k. \tag{481}$$

Noting that $\lambda = j\omega$, we see that z_k must be a coil having an inductance of a_k henries, and y_k a condenser with a capacitance equal to b_k farads.[1]

[1] Here it should be observed that the positiveness of the slope of $Z(j\omega)$ vs. ω ($\lambda = j\omega$) and the consequent separation property of the negative real λ^2 roots of P and Q insures the positiveness of the coefficients a_k and b_k and hence the physical realizability of the network. This may be seen by following an argument similar to that given by Cauer for the R, C case (*loc. cit.*, p. 363). If we let $Q = \theta\lambda$, then the continued fraction expansion takes the form

$$\lambda Z_{11}(\lambda) = a_1\lambda^2 + \frac{1|}{|b_1} + \frac{1}{|P'/\theta'},$$

and we see that

$$P' = P - \theta a_1\lambda^2;\ \theta' = \theta - P'b_1.$$

As functions of λ^2, we note that at zeros of P, P' has the same sign as θ (because the λ^2 roots are negative), while at zeros of θ it has the same sign as P. P', therefore, has the same number of roots as θ, and the roots of θ and P' separate each other. Similarly, at zeros of θ, θ' has the opposite sign to P', and at zeros of P' it has the

The network of Fig. 65 is, therefore, a physical realization for the reactance function (437), where

$$L_k = a_k;\ C_k = b_k;\ (k = 1, 2, 3, \cdots n). \tag{482}$$

If the expression (437) is first turned end-for-end, and the same process of division and inversion carried out, a continued fraction will result in which each term involves λ^{-1}. This may be written

$$Z_{11} = C_1^{-1}\lambda^{-1} + \cfrac{1}{L_1^{-1}\lambda^{-1} + \cfrac{1}{C_2^{-1}\lambda^{-1} + \cfrac{1}{L_2^{-1}\lambda^{-1} + \cdots + \cfrac{1}{C_n^{-1}\lambda^{-1} + \cfrac{1}{L_n^{-1}\lambda^{-1}}}}}}, \tag{483}$$

which is physically realized by means of the network of Fig. 66. This is the second alternative canonic form which Cauer gives.

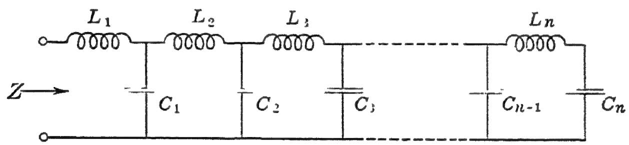


FIG. 65.—Cauer form for a driving-point reactance obtained by a continued fraction expansion of the reactance function.

In the special case which results when $Z_{11}(\lambda)$ has a zero at the origin, $\alpha_0 = 0$, so that the last term in (480) or the first term in (483) is absent. Then the condenser C_n in Fig. 65 or the condenser C_1 in Fig. 66 is absent, i.e., short-circuited.

In the special case for which $Z_{11}(\lambda)$ has a zero at infinity, $\alpha_{2n} = 0$, so that the first term in (480) or the last term in (483) is absent. Then the coil L_1 in Fig. 65 or the coil L_n in Fig. 66 is absent.

When $Z_{11}(\lambda)$ has zeros at the origin and at infinity, then L_1 and C_n in Fig. 65 or C_1 and L_n in Fig. 66 are both absent.

As an illustration of these methods let us consider again the reactance

---

same sign as θ. Thus θ' has one root less than P', and the roots of P' and θ' separate each other. Since the function P/θ is positive for $\lambda = 0$ (due to the positive slope of Z), the same is true for θ/P' as well as for P'/θ'. Hence P'/θ' is the same kind of function as P/θ, and a continuation of the argument leads to the same conclusions for P''/θ'' . . . etc., thus proving the positiveness of all the coefficients.

function (475) for the data (477) and $H = 5$. Multiplying out the factored form we find

$$Z_{11}(\lambda) = \frac{5\lambda^7 + 4.15\cdot 10^6\lambda + 9.45\cdot 10^{11}\lambda^3 + 5.51\cdot 10^{16}\lambda}{\lambda^6 + 5.6\cdot 10^5\lambda^4 + 7.84\cdot 10^{10}\lambda^2 + 2.3\cdot 10^{15}}. \qquad (484)$$

Beginning the division on the left we obtain the continued fraction

$$Z_{11}(\lambda) = 5\lambda + \frac{1|}{|0.74\cdot 10^{-6}\lambda} + \frac{1|}{|9\lambda} + \frac{1|}{|1.087\cdot 10^{-6}\lambda} + \frac{1|}{|6.5\lambda} + \frac{1|}{|2.68\cdot 10^{-6}\lambda} + \frac{1|}{|3.43\lambda}.$$

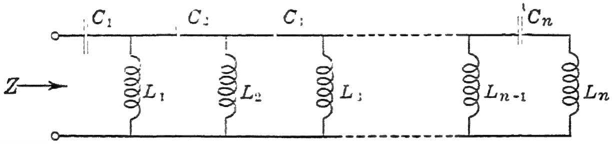


FIG. 66.—Alternative Cauer form for a driving-point reactance obtained by a continued fraction expansion of the reactance function.

The network is that shown in Fig. 67 where

$$L_1 = 5;\ L_2 = 9;\ L_3 = 6.5;\ L_4 = 3.43;$$
$$C_1 = 0.74\cdot 10^{-6};\ C_2 = 1.087\cdot 10^{-6};\ C_3 = 2.68\cdot 10^{-6}.$$

If we turn (484) end for end and carry out the dividing and inverting process we find

$$Z_{11}(\lambda) = \frac{1|}{|23.9^{-1}\lambda^{-1}} + \frac{1|}{|(0.706\cdot 10^{-6})^{-1}\lambda^{-1}} + \frac{1|}{|10.2^{-1}\lambda^{-1}} + \frac{1|}{|(0.227\cdot 10^{-6})^{-1}\lambda^{-1}} + \frac{1|}{|18.9^{-1}\lambda^{-1}} + \frac{1|}{|(0.0212\cdot 10^{-6})^{-1}\lambda^{-1}} + \frac{1|}{|139^{-1}\lambda^{-1}}.$$

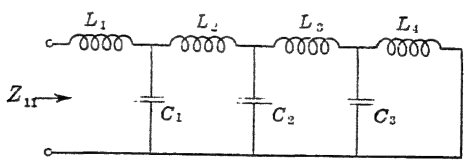


FIG. 67.—Cauer representation of the driving-point reactance of Fig. 60 according to the canonic form of Fig. 65.

The corresponding network is that shown in Fig. 68 where

$$L_1 = 23.9;\ L_2 = 10.2;\ L_3 = 18.9;\ L_4 = 139;$$
$$C_2 = 0.706\cdot 10^{-6};\ C_3 = 0.227\cdot 10^{-6};\ C_4 = 0.0212\cdot 10^{-6}.$$

7. Equivalence and reciprocity. By means of Foster's theorem and Cauer's extension thereof, we have found four fundamental equivalent network representations for a single reactance function each of which involves the least number of elements. By combining these methods in various ways, many other, equivalent networks may be found. For example, we may take any portion of either of Foster's networks and convert this into either of Cauer's and then replace it where it was removed. The result will still be equivalent but will have a different structure.

This brings us in contact with a very interesting and fruitful problem, namely that of finding all the networks equivalent to a given one. Although the above-mentioned combinations give us a large number of equivalents, these by no means represent the complete solution. Up to the present the problem is only partially solved, and we shall not go into detail here regarding its various aspects. We do wish to emphasize again, however, that any two driving-point reactances are equal when their poles and zeros coincide and, in addition, they are equal at one other frequency. Note that the networks are not equivalent if their reactance functions merely possess the same distribution of zeros and poles. Such networks are, however, referred to as **potentially equivalent** since they may be made equivalent merely by assigning the proper parameter values. Thus any two anti-resonant components are potentially equivalent. They will become equivalent if their parameter values are so chosen as to make their anti-resonant frequencies coincide and to make them equal at one additional point.

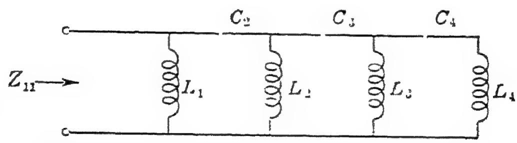


FIG. 68.—Alternative Cauer representation of the driving-point reactance of Fig. 60 according to the canonic form of Fig. 66.

In addition to equivalence there is another important mutual relationship between two-terminal networks. Consider, for example, the three functions Z_1, Z_2, and Z_3. If

$$Z_1Z_2 = Z_3^2, \tag{485}$$

then Z_2 is said to be the **inverse** or the **reciprocal** of Z_1 with respect to Z_3^2; or Z_1 is the reciprocal of Z_2 with respect to Z_3^2. A useful case results when Z_3 is a resistance, i.e., a positive real constant independent of frequency. In particular, if this constant has the value unity, then the impedances Z_1 and Z_2 are simply spoken of as being reciprocal or inverse.

When two non-dissipative two-terminal networks are so constituted

that the poles and zeros of their reactance functions mutually replace each other and their product equals the value R^2 at one other frequency, then it equals this value at all frequencies, i.e., the reactances are reciprocal with respect to the constant R^2. This may be seen to follow directly from Foster's theorem. Namely, if we regard the factored form (440a) as one of the reactance functions, say Z_1 with the factor H_1, then Z_2 will equal a factor H_2 multiplied by a quotient of factors which is the inverse of that for Z_1 because Z_2 is to have poles where Z_1 has zeros, and vice versa. The product of Z_1 and Z_2 will, therefore, equal H_1H_2. If $H_1H_2 = R^2$, then $Z_1Z_2 = R^2$. Since H_1H_2 depends only upon the values of Z_1 and Z_2 at one other frequency, our above statement is proved.

The necessary and sufficient condition for two reactances to be reciprocal is, therefore, that their zeros and poles shall mutually replace each other and that their product shall equal unity at one other frequency. If only their zeros and poles replace each other, then their product will equal a constant at all frequencies, but not necessarily unity. They will then be reciprocal with respect to this constant.

If the zeros and poles of a pair of driving-point reactance functions do not mutually replace each other except at the origin and at infinity, but the total number of internal zeros and poles is the same in both, then the corresponding networks are said to be **potentially reciprocal** because they can be made reciprocal with respect to any constant by a proper choice of parameter values. Thus a resonant and an anti-resonant component are potentially reciprocal networks. If their LC-products are made equal, the zero of one will coincide with the pole of the other, and their product will equal a constant at all frequencies. The reader should illustrate other cases for himself.

8. Bartlett's reciprocation theorem. Reciprocal structures play an important part in filter theory. There we are frequently faced with the problem of finding a network whose reactance function shall be the reciprocal of that for a given network with respect to an assigned constant. The solution to one form of this problem, which has been given by Bartlett,[1] will now be discussed.

Suppose we have given the network of Fig. 69 whose driving-point impedance is expressed by the following continued fraction

$$Z_1 = z_1 + \frac{1\,|}{|\,z_2^{-1}} + \frac{1\,|}{|\,z_3} + \frac{1\,|}{|\,z_4^{-1}} + \cdots + \frac{1\,|}{|\,z_{n-1}} + \frac{1\,|}{|\,z_n^{-1}}, \tag{486}$$

[1] A. C. Bartlett, "A Class of Equivalent Electrical Networks," Phil. Mag., Vol. 49, pp. 728–739, 1925. A more general discussion will be given in the next chapter.

and we wish to determine the reciprocal of this with respect to a constant R^2. According to (485) this reciprocal impedance is given by

$$Z_2 = \frac{R^2}{Z_1}, \tag{487}$$

which is the reciprocal of Z_1 multiplied by R^2. If we use (486) this becomes

$$Z_2 = \frac{1|}{\left(\frac{z_1}{R^2}\right)} + \frac{1|}{\left|\left(\frac{R^2}{z_2}\right)\right.} + \frac{1|}{\left|\left(\frac{z_3}{R^2}\right)\right.} + \frac{1|}{\left|\left(\frac{R^2}{z_4}\right)\right.} + \cdots + \frac{1|}{\left|\left(\frac{z_{n-1}}{R^2}\right)\right.} + \frac{1|}{\left|\left(\frac{R^2}{z_n}\right)\right.}, \tag{488}$$

because the factor R^2 alternately appears in the numerator and denominator of the succeeding fractions. This, however, we recognize as

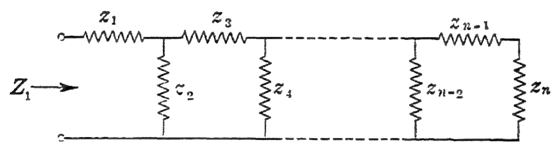


FIG. 69.—Two-terminal network whose impedance is given by eq. (486). The reciprocated structure, with respect to R^2, is shown in Fig. 70.

the driving-point impedance of the network shown in Fig. 70 which is the desired reciprocal.

A series of simple applications are illustrated in Fig. 71. The given networks are on the left and their reciprocals with respect to R^2 are on

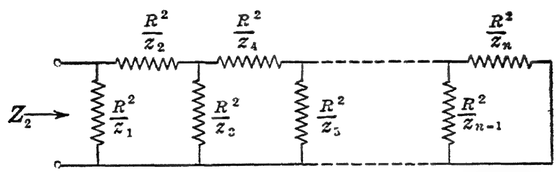


FIG. 70.—Reciprocal structure to that of Fig. 69 with respect to R^2.

the right. The reader should check the indicated parameter values as an exercise.

With any of these as component impedances in the given ladder network of Fig. 69, the reciprocal given by Fig. 70 may be set down by inspection. If the component impedances in the given ladder network are themselves ladder structures, the reciprocal with respect to R^2 is obtained by the same general process. This process Bartlett calls

reciprocation, i.e., Z_2 is found by *reciprocating* Z_1 *with respect to* R^2, or vice versa.

By means of the reciprocation theorem we are able to treat the general dissipative case provided the given network as well as its components are in the form of ladder structures.

A special case is of particular interest because of its usefulness in

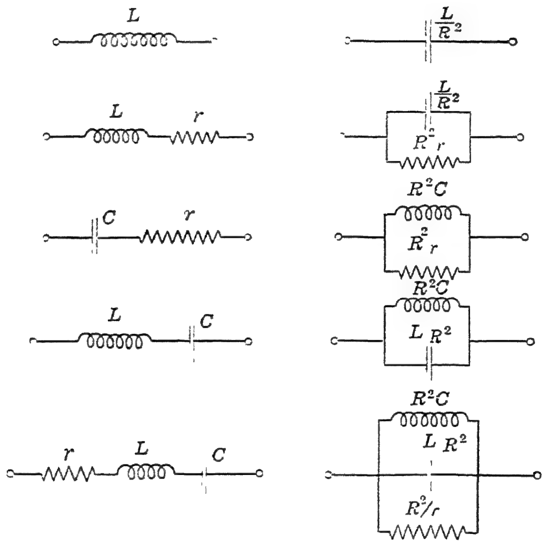


FIG. 71.—Simple element combinations and their reciprocals with respect to R^2.

connection with filter theory. For this the given network may, for example, be in the form of a number of components in series, as shown in Fig. 72. This has the same form as the ladder structure of Fig. 69 where all the shunt impedances except the last have become infinite,

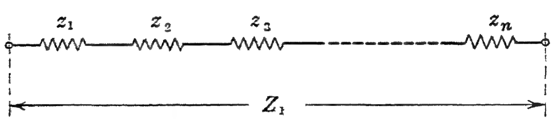


FIG. 72.—Special case of the ladder network of Fig. 69.

while the last shunt impedance is zero. The reciprocal with respect to R^2, according to Fig. 70, is, therefore, that shown in Fig. 73. Thus *when* Z_1 *is given as a number of component impedances in series, its reciprocal is a parallel combination of the reciprocals of the individual*

components, all with respect to the same constant R^2. Since Z_2 may as well be considered as the given network, the reverse process is also true. For example, a number of anti-resonant components in series, and the same number of resonant components in parallel, are each other's potentially inverse networks. If each anti-resonant component is the reciprocal with respect to R^2 of a corresponding resonant component, then the two resulting networks are each other's reciprocals with respect to R^2.

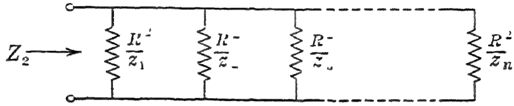


FIG. 73.—Reciprocal of the network of Fig. 72 with respect to R^2.

These networks are illustrated more explicitly in Fig. 74. Here we have

$$\left.\begin{aligned} Z_1 &= z_1 + z_3 + \cdots + z_{2n-1} \\ \text{and} \qquad \frac{1}{Z_2} &= \frac{1}{z_2} + \frac{1}{z_4} + \cdots + \frac{1}{z_{2n}} \end{aligned}\right\}. \tag{489}$$

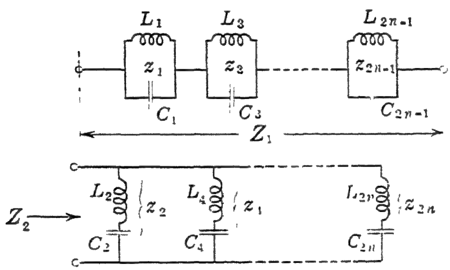


FIG. 74.—A pair of driving-point reactances which are potentially inverse.

From what has been said above it follows that we will have

$$Z_1Z_2 = R^2$$

if

$$z_1z_2 = z_3z_4 = \cdots = z_{2n-1}\, z_{2n} = R^2. \tag{490}$$

But

$$z_1 = \frac{jL_1\omega}{1 - L_1C_1\omega^2};\quad z_2 = \frac{1 - L_2C_2\omega^2}{jC_2\omega},$$

so that

$$z_1z_2 = \frac{L_1}{C_2} \cdot \frac{1 - L_2C_2\omega^2}{1 - L_1C_1\omega^2} = R^2$$

gives

$$\frac{L_1}{C_2} = \frac{L_2}{C_1} = R^2.$$

Hence, for the structures Z_1 and Z_2 of Fig. 74 to be reciprocal with respect to R^2, the parameters must fulfill the conditions

$$\frac{L_1}{C_2} = \frac{L_3}{C_4} = \cdots = \frac{L_{2n-1}}{C_{2n}} = \frac{L_2}{C_1} = \frac{L_4}{C_3} = \cdots = \frac{L_{2n}}{C_{2n-1}} = R^2. \quad (490a)$$

9. Extension of the reactance theorem to dissipative cases. Cauer[1] not only extended Foster's reactance theorem so as to enable us to obtain two additional canonic forms for driving-point reactance functions by means of continued fraction expansions, but also showed that the same general methods can also be used to obtain physical realizations for *dissipative* impedance functions when these correspond to networks containing *two* kinds of elements only, i.e., resistance and inductance only or resistance and capacitance only. Although such networks at present play a rather minor part in design problems as compared to the reactive networks, their use in the design of corrective networks and line balances warrants that we spend some time with them here.

For their discussion we begin as we did with the reactive case, namely with the formulation of the most general type of driving-point impedance function for networks containing R and L, or R and C only.

For the *R*, *L* case, the mesh and mutual impedances have the form

$$b_{ik} = R_{ik} + L_{ik}\lambda. \quad (491)$$

Hence the network determinant D is a polynomial in λ involving all powers from n to 0, while its minor B_{11} involves all powers from $n-1$ to 0, n being the number of independent meshes in the network. We may, therefore, write

$$Z_{11}(\lambda) = \frac{\alpha_n\lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0}{\beta_{n-1}\lambda^{n-1} + \beta_{n-2}\lambda^{n-2} + \cdots + \beta_1\lambda + \beta_0}. \quad (492)$$

The roots of these polynomials are negative real quantities. We shall denote the roots of the numerator and denominator respectively by

$$\left.\begin{matrix} \lambda_1, \lambda_3, \lambda_5, \cdots \lambda_{2n-1} \\ \lambda_2, \lambda_4, \lambda_6, \cdots \lambda_{2n-2} \end{matrix}\right\}. \quad (493)$$

On the negative real axis these roots possess the separation property

$$-\infty \le \lambda_{2n-1} < \lambda_{2n-2} < \cdots < \lambda_2 < \lambda_1 \le 0. \quad (493a)$$

In factored form the driving-point impedance function is

$$Z_{11}(\lambda) = H \cdot \frac{(\lambda - \lambda_1)(\lambda - \lambda_3) \cdots (\lambda - \lambda_{2n-1})}{(\lambda - \lambda_2)(\lambda - \lambda_4) \cdots (\lambda - \lambda_{2n-2})}. \quad (494)$$

[1] *Loc. cit.*, p. 268.

$Z_{11}(\lambda)$ may be expanded into partial fractions in the same way as the reactance function. We have first, however, to remove that portion which in (460) is denoted by $g(\lambda)$ and represents the series in ascending powers of λ. This consists in the present case only of a constant and a linear term in λ, since the poles of the function, including the one at infinity, are again simple. The constant term is that value which (492)

FIG. 75.

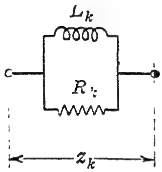


FIG. 76.

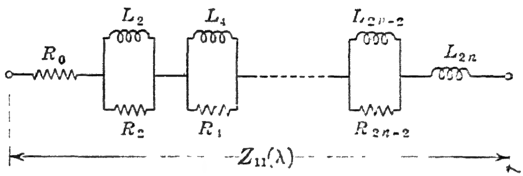


FIG. 75.—Network whose impedance function analytically represents a term in the partial fraction expansion (495a).

FIG. 76.—Network obtained from a partial fraction expansion of the driving-point impedance in the R,L case

has for $\lambda = 0$ and hence is α_0/β_0. The linear term is evaluated by letting $\lambda \to \infty$. Then

$$Z_{11}(\lambda) \to \frac{\alpha_n}{\beta_{n-1}}\lambda.$$

Dividing the constant and the linear terms out we have by (492)

$$Z_{11}(\lambda) = \frac{\alpha_0}{\beta_0} + \frac{\alpha_n}{\beta_{n-1}}\lambda + \frac{\alpha'_{n-1}\lambda^{n-1} + \alpha'_{n-2}\lambda^{n-2} + \cdots + \alpha_1'\lambda}{\beta_{n-1}\lambda^{n-1} + \beta_{n-2}\lambda^{n-2} + \cdots + \beta_0}, \tag{495}$$

where $\alpha'_{n-1} \ldots \alpha_1'$ are new coefficients formed by the dividing-out process. The partial fraction expansion of the impedance function for this case, therefore, takes the form

$$Z_{11}(\lambda) = \frac{\alpha_0}{\beta_0} + \frac{\alpha_n}{\beta_{n-1}}\lambda + \frac{k_2\lambda}{\lambda - \lambda_2} + \frac{k_4\lambda}{\lambda - \lambda_4} + \cdots + \frac{k_{2n-2}\lambda}{\lambda - \lambda_{2n-2}} \tag{495a}$$

where $k_2 \cdots k_{2n-2}$ are the residues of $Z_{11}(\lambda)$ in its poles $\lambda_2 \cdots \lambda_{2n-2}$. These are obtained in the same way as for the reactive case, and are

$$k_k = \left(\frac{dY_{11}(\lambda)}{d\lambda}\right)^{-1}_{\lambda=\lambda_k}; \ (k = 2, 4, \cdots 2n-2), \tag{496}$$

where $Y_{11}(\lambda)$ is the corresponding admittance function.

Considering now the component network of Fig. 75, we note that

$$z_k = \frac{R_k\lambda}{\lambda - \lambda_k}; \ \lambda_k = -\frac{R_k}{L_k}, \tag{497}$$

from which it becomes clear that the desired network is that shown in Fig. 76 where

$$R_0 = \frac{\alpha_0}{\beta_0};\ R_k = k_k;\ L_{2n} = \frac{\alpha_n}{\beta_{n-1}} = H. \tag{498}$$

Noting that

$$\left(\frac{dY_{11}}{d\lambda}\right)^{-1}_{\lambda=\lambda_k} = \left[\frac{Y_{11}}{\lambda - \lambda_k}\right]^{-1}_{\lambda=\lambda_k}, \tag{499}$$

we have with (494)

$$R_k = H \cdot \frac{(\lambda_k - \lambda_1)(\lambda_k - \lambda_3) \cdots (\lambda_k - \lambda_{2n-1})}{(\lambda_k-\lambda_2)(\lambda_k-\lambda_4) \cdots (\lambda_k-\lambda_{k-2})(\lambda_k-\lambda_{k+2}) \cdots (\lambda_k-\lambda_{2n-2})}. \tag{500}$$

The corresponding L_k's are found from (497)[1].

The network obtained in this way bears a close resemblance to that in the reactive case consisting of anti-resonant components in series.

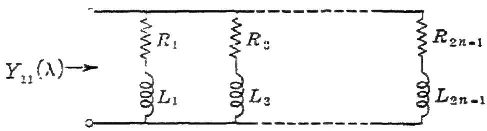


FIG. 77.—Network obtained from a partial fraction expansion of the driving-point admittance in the R,L case.

By expanding the reciprocal of (492) into partial fractions, we obtain the analogue to the reactive case consisting of resonant components in parallel. This expansion is

$$Y_{11}(\lambda) = \frac{k_1}{\lambda - \lambda_1} + \frac{k_3}{\lambda - \lambda_3} + \cdots + \frac{k_{2n-1}}{\lambda - \lambda_{2n-1}}. \tag{501}$$

Since $Y_{11}(\lambda)$ vanishes at infinity, $g(\lambda)$ does so also. Being a *rational integral* function of λ, $g(\lambda)$ must, therefore, vanish identically. Thus the partial fraction expansion of $Y_{11}(\lambda)$ is given by the terms in (501) alone.

For the residues of $Y_{11}(\lambda)$ in its poles we have

$$k_k = \left(\frac{dZ_{11}(\lambda)}{d\lambda}\right)_{\lambda=\lambda_k};\ (k = 1, 3, \cdots 2n - 1). \tag{502}$$

The network represented by (501) is shown in Fig. 77, for which

$$\frac{1}{L_k} = k_k;\ \lambda_k = -\frac{R_k}{L_k}. \tag{503}$$

[1] The positiveness of the residues is again a necessary condition for physical realizability. In addition the roots λ_k must all be negative real.

Since

$$\left(\frac{dZ_{11}}{d\lambda}\right)^{-1}_{\lambda=\lambda_k} = \left[\frac{Z_{11}}{\lambda - \lambda_k}\right]^{-1}_{\lambda=\lambda_k}, \tag{504}$$

we have

$$L_k = H \cdot \frac{(\lambda_k - \lambda_1) \cdots (\lambda_k - \lambda_{k-2})(\lambda_k - \lambda_{k+2}) \cdots (\lambda_k - \lambda_{2n-1})}{(\lambda_k - \lambda_2)(\lambda_k - \lambda_4) \cdots (\lambda_k - \lambda_{2n-2})}. \tag{505}$$

In addition to these fundamental forms, two additional ones may be obtained by continued fraction expansions of (492). Beginning the division on the left we get

$$Z_{11}(\lambda) = L_1\lambda + \frac{1\,|}{|R_1^{-1}} + \frac{1\,|}{|L_2\lambda} + \frac{1\,|}{|R_2^{-1}} + \cdots + \frac{1\,|}{|L_n\lambda} + \frac{1}{|R_n^{-1}}, \tag{506}$$

for which the network representation is shown in Fig. 78*a*.

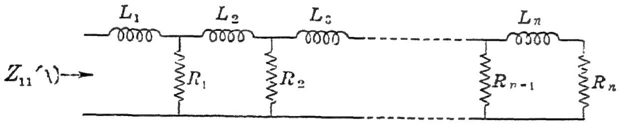


FIG. 78*a*.—Network obtained from a continued fraction expansion of the driving-point impedance in the R,L case.

Beginning the division on the right we have

$$Z_{11}(\lambda) = R_1 + \frac{1\,|}{|L_1^{-1}\lambda^{-1}} + \frac{1\,|}{|R_2} + \frac{1\,|}{|L_2^{-1}\lambda^{-1}} + \cdots + \frac{1\,|}{|R_n} + \frac{1}{|L_n^{-1}\lambda^{-1}}, \tag{506a}$$

which leads to the network of Fig. 78*b*.

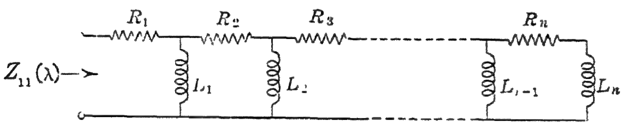


FIG. 78*b*.—Alternative network obtained from a continued fraction expansion of the driving-point impedance in the R,L case.

For the *R, C* case, the mesh and mutual impedances are

$$b_{ik} = R_{ik} + S_{ik}\lambda^{-1}. \tag{507}$$

Hence $\lambda^n D$ is a polynomial in λ ranging from λ^n to λ^0, and $\lambda^n B_{11}$ involves powers of λ from n to 1. The driving-point impedance function, therefore, has the form

$$Z_{11}(\lambda) = \frac{\alpha_n\lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0}{\beta_n\lambda^n + \beta_{n-1}\lambda^{n-1} + \cdots + \beta_2\lambda^2 + \beta_1\lambda}. \tag{508}$$

Denoting the roots of the numerator and denominator respectively by

$$\left.\begin{array}{l}\lambda_1,\ \lambda_3,\ \lambda_5,\ \cdots\ \lambda_{2n-1}\\ \lambda_0,\ \lambda_2,\ \lambda_4,\ \cdots\ \lambda_{2n-2}\end{array}\right\}, \tag{509}$$

and

their separation property is in this case expressed by

$$-\infty \leq \lambda_{2n-1} < \lambda_{2n-2} < \cdots < \lambda_2 < \lambda_1 < \lambda_0 = 0. \tag{509a}$$

The factored form for the impedance is

$$Z_{11}(\lambda) = H \cdot \frac{(\lambda - \lambda_1)(\lambda - \lambda_3)\ \cdots\ (\lambda - \lambda_{2n-1})}{\lambda(\lambda - \lambda_2)(\lambda - \lambda_4)\ \cdots\ (\lambda - \lambda_{2n-2})}. \tag{510}$$

The partial fraction expansion of $Z_{11}(\lambda)$ is

$$Z_{11}(\lambda) = \frac{k_0}{\lambda} + \frac{k_2}{\lambda - \lambda_2} + \cdots + \frac{k_{2n-2}}{\lambda - \lambda_{2n-2}} + \frac{\alpha_n}{\beta_n}, \tag{510a}$$

in which

$$k_k = \left(\frac{dY_{11}(\lambda)}{d\lambda}\right)^{-1}_{\lambda=\lambda_k};\ (k = 0, 2, 4, \cdots 2n - 2). \tag{511}$$

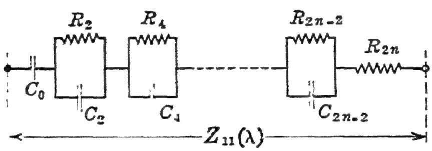


FIG. 79.—Network realization for the R,C case obtained from a partial fraction expansion of the driving-point impedance function.

The corresponding network is shown in Fig. 79 with the parameter values

$$\left.\begin{array}{l}C_k = \left[\dfrac{Y_{11}(\lambda)}{\lambda - \lambda_k}\right]_{\lambda=\lambda_k};\ (k = 0, 2, 4, \cdots 2n - 2)\\[2ex] \lambda_k = -\dfrac{1}{R_kC_k};\ R_{2n} = \dfrac{\alpha_n}{\beta_n} = H\end{array}\right\}. \tag{512}$$

Expanding the reciprocal of (508) or (510) into partial fractions we get

$$Y_{11}(\lambda) = \frac{k_1\lambda}{\lambda - \lambda_1} + \frac{k_3\lambda}{\lambda - \lambda_3} + \cdots + \frac{k_{2n-1}\lambda}{\lambda - \lambda_{2n-1}}, \tag{513}$$

in which

$$k_k = \frac{1}{\lambda_k}\left(\frac{dZ_{11}}{d\lambda}\right)^{-1}_{\lambda=\lambda_k} = \left[\frac{\lambda \cdot Z_{11}(\lambda)}{\lambda - \lambda_k}\right]^{-1}_{\lambda=\lambda_k};\ (k = 1, 3, \cdots 2n - 1). \tag{514}$$

The corresponding network is that given in Fig. 80 for which

$$\frac{1}{R_k} = k_k; \ \lambda_k = -\frac{1}{R_k C_k}. \tag{515}$$

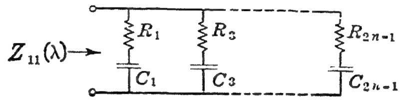


FIG. 80.—Network realization for the R,C case obtained from a partial fraction expansion of the driving-point admittance function.

Beginning at the left, the continued fraction expansion of (508) gives

$$Z_{11}(\lambda) = R_1 + \frac{1\,|}{|C_1\lambda} + \frac{1\,|}{|R_2} + \frac{1\,|}{|C_2\lambda} + \cdots + \frac{1\,|}{|R_n} + \frac{1}{C_n\lambda}, \tag{516}$$

which is realized by the network of Fig. 81a.

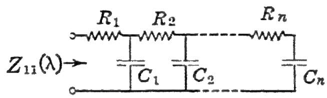


FIG. 81a.—Network realization in the R,C case obtained from a continued fraction expansion of the driving-point impedance function.

Beginning the continued fraction expansion of (508) on the right, we have

$$Z_{11}(\lambda) = C_1^{-1}\lambda^{-1} + \frac{1\,|}{|R_1^{-1}} + \frac{1\,|}{|C_2^{-1}\lambda^{-1}} + \frac{1\,|}{|R_2^{-1}} + \cdots + \frac{1\,|}{|C_n^{-1}\lambda^{-1}} + \frac{1}{R_n^{-1}}, \tag{516a}$$

for which the corresponding network is shown in Fig. 81b.

The above are for the general cases where each of the n meshes independently contains R and L or R and C. Special examples for which this is not true are treated by the same general processes. We shall conclude this discussion with an illustrative example for one of the dissipative cases.

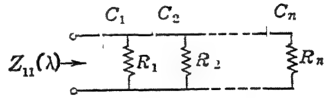


FIG. 81b.—Network realization in the R,C case obtained from an alternative continued fraction expansion of the driving-point impedance.

Suppose we wish to find a network whose driving-point impedance function $Z_{11}(\lambda)$ has poles for $\lambda = 0, -2, -4$, and zeros for $\lambda = -1, -3$. Since the origin is a pole, the network must be of the R, C type. Its impedance function will have the form

$$Z_{11}(\lambda) = H \cdot \frac{(\lambda+1)(\lambda+3)}{\lambda(\lambda+2)(\lambda+4)} = H \cdot \frac{\lambda^2+4\lambda+3}{\lambda^3+6\lambda^2+8\lambda}.$$

For $\lambda = j\omega$ this is

$$Z_{11}(j\omega) = H \cdot \frac{(2\omega^4 + 14\omega^2) - j(\omega^5 + 13\omega^3 + 24\omega)}{\omega^6 + 2\omega^4 + 64\omega^2}.$$

We see that the real part is an even function while the imaginary part is odd. This holds generally.[1] Moreover, *for* R, C *networks the imaginary part is negative while the real part is a decreasing function,* and *for* R, L *networks the imaginary part is positive while the real part is an increasing function.*

For the determination of H let us say that the real part of $Z_{11}(j\omega)$ shall be equal to unity at the origin. This gives

$$H \cdot \frac{14}{64} = 1; \; H = \frac{32}{7}.$$

For the realization of this function in the form of Fig. 79, we apply (512) where

$$Y_{11}(\lambda) = \frac{7}{32} \cdot \frac{\lambda(\lambda+2)(\lambda+4)}{(\lambda+1)(\lambda+3)},$$

and
$$\lambda_0 = 0, \lambda_2 = -2, \lambda_4 = -4.$$

Thus we get for the capacitances

$$\left.\begin{aligned} C_0 &= \frac{7}{32} \cdot \frac{2 \cdot 4}{1 \cdot 3} = \frac{7}{12} \\ C_2 &= \frac{7}{32} \cdot \frac{2 \cdot 2}{1 \cdot 1} = \frac{7}{8} \\ C_4 &= \frac{7}{32} \cdot \frac{4 \cdot 2}{3 \cdot 1} = \frac{7}{12} \end{aligned}\right\}.$$

The resistances which are associated with C_2 and C_4 are found by using

$$\lambda_k = -\frac{1}{R_k C_k} \text{ for } k = 2, 4,$$

thus

$$R_2 = \frac{4}{7}; \; R_4 = \frac{3}{7}.$$

The R_{2n} of Fig. 79 is zero because $Z_{11}(\lambda)$ vanishes for $\lambda \to \infty$. This is

[1] This may be seen from the fact that any driving-point impedance is given by $Z_{11} = D/B_{11}$ where D is the network determinant and B_{11} the minor of its first row and column. Since the elements of D are $b_{ik} = R_{ik} + j\omega(L_{ik} - S_{ik}/\omega^2)$, and both D and B_{11} consist of sums of products of these elements, the real parts of the determinant and its minor are even functions of ω while the imaginary parts are odd. Thus the real part of their quotient must be even and the imaginary part odd, so that this property is evident in both Z_{11} and its reciprocal Y_{11}. The same argument holds with respect to any of the transfer impedances or admittances.

the same as to say that $\alpha_{2n} = 0$. The reader may draw the resulting network himself.

In order to get the network realization of Fig. 80, we consider the partial fraction expansion of $Y_{11}(\lambda)$ in the form (513). In this case $Z_{11}(\lambda)$ has a zero at infinity so that $Y_{11}(\lambda)$ has a pole there. The partial fraction expansion has the form

$$Y_{11}(\lambda) = \frac{k_1\lambda}{\lambda - \lambda_1} + \frac{k_3\lambda}{\lambda - \lambda_3} + k_5\lambda.$$

The residues at the poles λ_1 and λ_3 may be evaluated by (514). With (515) this gives

$$\left.\begin{aligned} R_1 &= \frac{32}{7} \cdot \frac{2}{1 \cdot 3} = \frac{64}{21};\ C_1 = \frac{21}{64} \\ R_3 &= \frac{32}{7} \cdot \frac{2}{1 \cdot 1} = \frac{64}{7};\ C_3 = \frac{7}{192} \end{aligned}\right\}.$$

The pole at infinity is represented by a condenser alone. By inspection we see that $k_5 = 7/32$. Hence we have

$$R_5 = 0;\ C_5 = \frac{7}{32},$$

which completes the realization according to Fig. 80.

For the continued fraction expansions of $Z_{11}(\lambda)$ in this case we find if we begin at the left

$$Z_{11}(\lambda) = \frac{1|}{|\frac{7}{32}\lambda} + \frac{1|}{|\frac{16}{7}} + \frac{1|}{|\frac{28}{96}\lambda} + \frac{1|}{|\frac{48}{7}} + \frac{1|}{|\frac{7}{96}\lambda},$$

which leads to the realization according to Fig. 81a with

$$R_1 = 0;\ R_2 = \frac{16}{7};\ R_3 = \frac{48}{7}$$

$$C_1 = \frac{7}{32};\ C_2 = \frac{28}{96};\ C_3 = \frac{7}{96}.$$

Beginning at the right we find

$$Z_{11}(\lambda) = \left(\frac{7}{12}\right)^{-1}\lambda^{-1} + \frac{1|}{|1} + \frac{1|}{|\left(\frac{11}{28}\right)^{-1}\lambda^{-1}} + \frac{1|}{|\left(\frac{12}{121}\right)^{-1}} + \frac{1|}{|\left(\frac{77}{24}\right)^{-1}\lambda^{-1}},$$

for which Fig. 81b illustrates the network where

$$C_1 = \frac{7}{12};\ C_2 = \frac{11}{28};\ C_3 = \frac{77}{24};$$

$$R_1 = 1;\ R_2 = \frac{12}{121};\ R_3 = 0$$

It may be of interest to the reader to know that the general dissipative case where $Z_{11}(\lambda)$ is the driving-point impedance of a network containing *all three kinds* of elements (R, L, and C) has been solved by O. Brune[1] who shows that the necessary and sufficient condition for physical realization is stated by saying that *the real part of* $Z_{11}(\lambda)$ *should be positive for positive values of the real part of* $\lambda = \gamma + j\omega$, *i.e.*, $Re\ [Z_{11}(\lambda)] \geq 0$ for $\gamma \geq 0$. This condition may also be stated in terms of the boundary ($\lambda = j\omega$; $\gamma = 0$) alone, namely: For $\gamma = 0$, $Re\ [Z_{11}(\lambda)] \geq 0$, *and in addition the residues at those poles which occur on the boundary (imaginary axis) must be positive.* When $Z_{11}(\lambda)$ satisfies these conditions it is said to be a **positive real function.** In the reactive case the real part of $Z_{11}(\lambda)$ on the boundary is zero, and the condition reduces to the positiveness of the residues only, as has been pointed out above. So far the solution to the general dissipative case has been of little utility in network design, and we shall, therefore, not consume space with a discussion of it here.

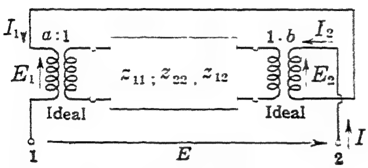


FIG. 82.—Reduction of a four-terminal to a two-terminal network for the determination of the conditions for physical realizability.

10. The reactance theorem extended to a two-terminal pair.[2] As pointed out in the previous chapter the two-terminal pair is characterized by three functions, for example, by z_{11}, z_{22}, z_{12}; or y_{11}, y_{22}, y_{12}; etc. In the non-dissipative case these are, of course, reactance (susceptance) functions similar to the $Z_{11}(\lambda)$ considered in the first six sections above. When three such reactance functions are specified to be, let us say, the *z-system* of a two-terminal pair, the question arises: Under what conditions can such a system of functions lead to a physical structure, and what procedure may be followed in order to realize it?

The answer to the first part of this question is given in a remarkably simple fashion by a method of reasoning suggested by Brune.[3] The procedure is essentially to reduce the problem to that of a driving-

[1] "Synthesis of a Finite Two-Terminal Network Whose Driving-Point Impedance is a Prescribed Function of Frequency," Jour. of Math. and Phys., Vol. X, No. 3, 1931, p. 191.

[2] An extension of Foster's reactance theorem to the general case of n pairs of terminals is given by W. Cauer in his paper entitled "Ein Reaktanztheorem," Sitzungsberichte d. Preuss. Akad. d. Wiss. Phys.-Math. Klasse Vol. XXX, 1931. The present section discusses in detail only the case of two pairs of terminals.

[3] The following device is the extension of an idea originally due to Cauer and employed in his paper "Vierpole" (E.N.T. Vol. 6. No. 7, 1929 on page 279) for the purpose of showing that any symmetrical two-terminal pair may be realized in the lattice form.

point reactance by the artifice illustrated in Fig. 82. The given two-terminal pair is characterized by the system z_{11}, z_{22}, z_{12}. With the ideal transformers we have

$$\left.\begin{aligned} E_1 &= a^2 z_{11} I_1 + ab z_{12} I_2 \\ E_2 &= ab z_{21} I_1 + b^2 z_{22} I_2 \end{aligned}\right\}. \tag{517}$$

Denoting the voltage rise from terminal 1 to 2 by E and the current entering 2 by I, we have

$$\left.\begin{aligned} E &= E_1 + E_2 \\ I &= I_1 = I_2 \end{aligned}\right\},$$

so that the addition of (517) gives

$$E = (a^2 z_{11} + 2ab z_{12} + b^2 z_{22}) I. \tag{518}$$

The reactance at the terminals $1 - 2$ is, therefore,

$$Z(\lambda) = \frac{E}{I} = z_{11} a^2 + 2 z_{12} ab + z_{22} b^2. \tag{518a}$$

This may be looked upon as a *quadratic form* in the variables a and b which are the ratios of the ideal transformers. If we denote the residue of $Z(\lambda)$ at a pole corresponding to $\lambda = \lambda_\nu$ by k_ν, then it follows from (518a) that

$$k_\nu = k_{11}^{(\nu)} a^2 + 2 k_{12}^{(\nu)} ab + k_{22}^{(\nu)} b^2, \tag{518b}$$

where $k_{11}^{(\nu)}$ $k_{12}^{(\nu)}$ $k_{22}^{(\nu)}$ are the residues of z_{11}, z_{12}, and z_{22} at $\lambda = \lambda_\nu$. Physical realizability now demands that the residue k_ν be positive for all possible values of a and b. The necessary and sufficient condition for this is that the determinant of the quadratic form (518b) be positive.[1] Hence

$$\begin{vmatrix} k_{11}^{(\nu)} & k_{12}^{(\nu)} \\ k_{21}^{(\nu)} & k_{22}^{(\nu)} \end{vmatrix} = k_{11}^{(\nu)} k_{22}^{(\nu)} - (k_{12}^{(\nu)})^2 \geq 0 \tag{519}$$

becomes the desired condition for the physical realization of the given reactive two-terminal pair.[2] The same type of reasoning can be carried

[1] See Section 2 of Chapter VI. It is understood, of course, that the residues k_{12}^{ν} and k_{22}^{ν} are *real* and *positive*.

[2] In the dissipative case the condition for physical realizability becomes $r_{11}r_{22} - r_{12}^2 \geq 0$ for $\gamma \geq 0$, $(\lambda = \gamma + j\omega)$ where r_{11}, r_{22}, and r_{12} are the real parts of z_{11}, z_{22}, and z_{12} respectively. A *z-matrix* satisfying this condition is called **positive real.** As in the two-terminal case this condition may be restricted to the boundary $\lambda = j\omega$ when supplemented by the further conditions (519) for poles on the boundary. An alternative derivation of these necessary conditions for physical realizability in the dissipative case is given by C. Gewertz in his paper entitled "Synthesis of a Finite, Four-Terminal Network from its Prescribed Driving-point Functions and Transfer Function," Jl. of Math. & Phys., Vol. 12, 1932–33, pp. 1–257 who also shows the sufficiency of these conditions.

out in terms of the *y-system* by placing the input and output ends of the network of Fig. 82 in *parallel* instead of in series. Then the condition (519) is again arrived at where $k_{11}^{(\nu)}$, $k_{22}^{(\nu)}$, and $k_{12}^{(\nu)}$ are understood to be the residues of y_{11}, y_{22}, and y_{12} respectively. This corresponding condition is equivalent to the one derived for the *z-system* and hence states nothing new.

A possible physical realization for the network whose *z-system* satisfies the condition (519) at all of the poles of its reactances is easily given. The first step is to expand the reactances z_{11}, z_{22}, and z_{12} into partial fractions similar to (461*a*), thus

$$\left.\begin{aligned} z_{11}(\lambda) &= \frac{k_{11}^{(0)}}{\lambda} + \frac{2k_{11}^{(2)}\lambda}{\lambda^2 - \lambda_2^2} + \cdots + \frac{2k_{11}^{(2n-2)}\lambda}{\lambda^2 - \lambda^2_{2n-2}} + k_{11}^{(2n)}\lambda \\ z_{22}(\lambda) &= \frac{k_{22}^{(0)}}{\lambda} + \frac{2k_{22}^{(2)}\lambda}{\lambda^2 - \lambda_2^2} + \cdots + \frac{2k_{22}^{(2n-2)}\lambda}{\lambda^2 - \lambda^2_{2n-2}} + k_{22}^{(2n)}\lambda \\ z_{12}(\lambda) &= \frac{k_{12}^{(0)}}{\lambda} + \frac{2k_{12}^{(2)}\lambda}{\lambda^2 - \lambda_2^2} + \cdots + \frac{2k_{12}^{(2n-2)}\lambda}{\lambda^2 - \lambda^2_{2n-2}} + k_{12}^{(2n)}\lambda \end{aligned}\right\}. \quad (520)$$

Here it may be well to point out that a pole of z_{11} (or of z_{22}) may not be contained in z_{12} or z_{22} (resp. z_{11}). In this case the corresponding residue $k_{12}^{(\nu)}$ or $k_{22}^{(\nu)}$ (resp. $k_{11}^{(\nu)}$) becomes zero. Since a pole contained in z_{12} must also be contained in both z_{11} and z_{22}, the condition (519) can still be fulfilled without requiring that the three reactances have coincident poles throughout.

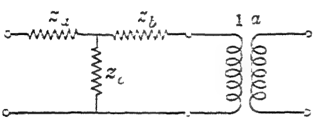


FIG. 83 —Component network for the synthesis of a reactive two-terminal pair according to the partial fraction expansions (520).

Let us designate the reactances of the first, second, third, etc., columns in (520) as *component z-systems*. If a network representation is found for each component *z-system*, then the series connection of these component networks in the manner illustrated by Fig. 37 (page 151) evidently represents the desired physical realization. Since the reactances in any component *z-system* bear constant (positive real) ratios to one another, and their coefficients satisfy the condition (519), each component network may be realized in the form of a T-network plus an ideal transformer as illustrated in Fig. 83. Writing the component *z-system* as

$$k_{11}f(\lambda),\ k_{22}f(\lambda),\ k_{12}f(\lambda),$$

and the reactances of the component T as

$$z_a = s_a f(\lambda),\ z_b = s_b f(\lambda),\ z_c = s_c f(\lambda),$$

we must have

$$\left.\begin{aligned} k_{11} &= s_a + s_c \\ k_{12} &= as_c \\ k_{22} &= a^2(s_b + s_c) \end{aligned}\right\}, \tag{521}$$

from which

$$\left.\begin{aligned} s_a &= k_{11} - \frac{k_{12}}{a} \\ s_b &= \frac{k_{22}}{a^2} - \frac{k_{12}}{a} \\ s_c &= \frac{k_{12}}{a} \end{aligned}\right\}. \tag{521a}$$

Since these must be positive, the ideal transformer ratio a is restricted to the range of values

$$\frac{k_{12}}{k_{11}} \leq a \leq \frac{k_{22}}{k_{12}}, \tag{522}$$

but is otherwise arbitrary. The lower limit of a makes $s_a = 0$ while the upper limit results in $s_b = 0$. The condition $k_{11}k_{22} - k_{12}^2 \geq 0$ insures positive values in any case as the reader may readily verify. The function $f(\lambda)$ represents a condenser, an anti-resonant component, or a coil, so that the final form of the component network is thus established. If $k_{11} > k_{12}$ *and* $k_{22} > k_{12}$, the range (522) includes the value *unity* so that the ideal transformer may be dispensed with.

An alternative method is to expand the *y-system* in partial fractions, realize each *component y-system* in the form of a π-network plus an ideal transformer, and then connect these component networks in parallel. These two methods of realization for the four-terminal network are analogous to Foster's forms for the two-terminal one.

PROBLEMS TO CHAPTER V

5-1. The driving-point reactance function illustrated in Fig. 55 has the following critical frequencies: $\omega_1 = 2000$, $\omega_2 = 4000$, $\omega_3 = 6000$, $\omega_4 = 7000$, $\omega_5 = 7500$. The residue in the pole at the origin is equal to 100. Determine the two Foster and the two Cauer networks which physically realize this function. Sketch a number of alternative potentially equivalent structures.

5-2. Determine the four canonic forms for the inverse reactance of that specified in Problem 5-1. Sketch additional potentially inverse structures.

5-3. A network has the following inductance and elastance matrices:

$$\begin{Vmatrix} L_1 & 0 & -L_1 & 0 & 0 \\ 0 & L_2 & -L_2 & 0 & 0 \\ -L_1 & -L_2 & (L_1 + L_2) & 0 & 0 \\ 0 & 0 & 0 & L_3 & 0 \\ 0 & 0 & 0 & 0 & L_4 \end{Vmatrix},$$

$$\left\| \begin{array}{ccccc} S_1 & -S_1 & 0 & 0 & 0 \\ -S_1 & S_1 & 0 & 0 & 0 \\ 0 & 0 & (S_2 + S_3) & -S_2 & -S_3 \\ 0 & 0 & -S_2 & S_2 & 0 \\ 0 & 0 & -S_3 & 0 & (S_3 + S_4) \end{array} \right\|.$$

Sketch the network. If the driving point lies on that part of the contour of mesh 1 which is not common to any other mesh, sketch the distribution of zeros and poles for the driving-point reactance function and sketch the four canonic potentially equivalent as well as the potentially inverse networks What is the least number of coils and condensers by means of which the network may be realized? Write down the analytic form of the driving-point reactance function.

5-4. A network whose driving point lies on that part of the contour of mesh 1 not common to the others has the following inductance and elastance matrices:

$$\left\| \begin{array}{ccc} (L_1 + L_2 + L_3) & -L_2 & -L_3 \\ -L_2 & (L_2 + L_4) & 0 \\ -L_3 & 0 & L_3 \end{array} \right\|; \quad \left\| \begin{array}{ccc} (S_1 + S_2 + S_3) & 0 & -S_2 \\ 0 & S_4 & 0 \\ -S_2 & 0 & (S_2 + S_5) \end{array} \right\|.$$

Determine the network and its four potentially equivalent canonic forms. How many superfluous elements are contained in the original network?

5-5. A driving-point reactance is to have zeros at the frequencies: $\omega_1 = 2500$, $\omega_3 = 4900$, and $\omega_5 = 6400$. The corresponding susceptance should have a slope equal to 10^{-4} mho per radian per second at each of these frequencies. Determine the network realizing these characteristics, and find the frequencies corresponding to the poles of the reactance function.

5-6. A network has the following resistance, inductance, and elastance matrices, respectively:

$$\mathsf{R} = \left\| \begin{array}{ccc} 5 & -4 & 0 \\ -4 & 14 & -10 \\ 0 & -10 & 17 \end{array} \right\|; \quad \mathsf{L} = 10^{-3} \cdot \left\| \begin{array}{ccc} 2 & 0 & 0 \\ 0 & 13 & -9 \\ 0 & -9 & 9 \end{array} \right\|;$$

$$\mathsf{S} = 10^{6} \cdot \left\| \begin{array}{ccc} 5 & -5 & 0 \\ -5 & 8 & 0 \\ 0 & 0 & 6 \end{array} \right\|.$$

The excitation lies on the contour of mesh 1 exclusively. Reciprocate the driving-point impedance so as to determine the inverse two-terminal network with respect to the real constant value 100.

5-7. It is desired to find a network realization for a dissipative driving-point impedance function having zeros corresponding to $\lambda = -10, -20, -30$, and poles corresponding to $\lambda = -15, -25$. The imaginary part should approach the value $10^{-3}\omega$ for high frequencies. What two kinds of elements does the network consist of? What is the value of the real part for $\omega = 0$, and is it an increasing or a decreasing function with frequency? Find the four canonic forms for this network.

5-8. Find the inverse of each of the four network realizations in Problem 5-7 with respect to a resistance of 1000 ohms.

5-9. It is desired to design a dissipative network whose driving-point impedance function shall simulate the characteristic impedance of a long, uniform transmission line. Show that if the line parameters are such that $\frac{R}{L} < \frac{G}{C}$ the simulative network

contains only resistance and inductance, while if $\frac{R}{L} > \frac{G}{C}$ the network consists only of resistance and capacitance elements. What two kinds of elements will be involved in practical lines and cables?

5–10. Utilizing the results of Problem 3–3, design an impedance simulative network for the open line whose data are specified in Problem 3–1, assuming (*a*) that the network involves only one mesh; (*b*) that it involves two meshes; and (*c*) that it involves three meshes. In part (*a*) consider both resistance and capacitance to be present; in (*b*) consider the added mesh to introduce an additional resistance alone, or an additional capacitance alone, or both. Do the same for part (*c*). In (*b*) and (*c*) note the improvement afforded by the added elements on the simulative ability of the real and imaginary parts of the resulting impedance function. The design procedure may be carried out by deriving in each case the analytic expressions for the real and imaginary parts and determining their parameters so as to result in the best average approximation to the curves for the characteristic impedance function of the line over the specified frequency range.

5–11. Repeat Problem 5–10 for the cable data given in Problem 3–2, for which the characteristic impedance function was determined in Problem 3–4.

5–12. The design of a symmetrical low-pass filter (see section 8, Chapter X) leads to the following open-circuit input and transfer reactance functions

$$z_{11} = \frac{0.577536 - 3.3598\,x^2 + 2.765\,x^4}{2jx\,(0.64 - x^2)\,(1.44 - x^2)},$$

$$z_{12} = \frac{-0.577536 + 0.6498\,x^2 - 0.055\,x^4}{2jx\,(0.64 - x^2)\,(1.44 - x^2)},$$

in which the variable $x = \omega/\omega_1$ with ω_1 as the theoretical cut-off frequency. Find a network realization for this two-terminal pair according to the extension of the Foster method discussed in section 10. Alternatively find a network realization by proceeding from the corresponding susceptance functions.

CHAPTER VI

ENERGY FUNCTIONS AND LINEAR NETWORK TRANSFORMATIONS

1. General remarks. The methods of the preceding chapter have shown that many networks exist which give rise to the same behavior with respect to one pair of terminals. This brings us into contact with a most interesting as well as useful collateral problem, namely, that of finding all equivalent networks when one is given. Foster's and Cauer's theorems give a partial solution to this question with regard to driving-point impedances. They by no means exhaust the situation even for two terminals. For one thing, they are limited to the treatment of networks involving only two kinds of network elements, namely R, C; R, L; and L, C networks. Furthermore, they give only such equivalents which fall under one of the canonic forms or series and parallel combinations of these. The question regarding the equivalence of networks with respect to more than one pair of terminals is completely outside the scope of the above theorems.

The problem of finding inverse networks, which is closely associated with the problem of equivalents, is partially solved by Bartlett's principle of reciprocation, but the general case is here also left in an unsolved state.

In the present chapter we wish to discuss a method of network transformation which, although by no means constituting a complete solution to this problem, will give us a more general conception of the principles involved and hence lead to a better understanding of the significance of the problem as a whole. Before proceeding with his method it is necessary to point out certain energy relations and their bearing upon the equilibrium conditions in linear networks.

2. Lagrange's equations for the dynamic equilibrium of linear systems. The formulation of the dynamic equilibrium conditions for a linear network was given in Chapter IV of Volume I by making use of Kirchhoff's emf equations. We wish here to show that the same conditions can be arrived at from energy considerations by the application of Lagrange's equations. Instead of proceeding from these, we believe that the engineering student will derive a better understanding of the matter if we start from more familiar network relations and lead up to Lagrange's equations as an ultimate goal.

With the same notation regarding mesh charges and currents as well as mesh and mutual parameters as have been used throughout the first volume, we may write the following equations

$$\left.\begin{array}{l} L_{11}\dot{q}_1 + L_{12}\dot{q}_2 + \cdots + L_{1n}\dot{q}_n = \varphi_1 \\ L_{21}\dot{q}_1 + L_{22}\dot{q}_2 + \cdots + L_{2n}\dot{q}_n = \varphi_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ L_{n1}\dot{q}_1 + L_{n2}\dot{q}_2 + \cdots + L_{nn}\dot{q}_n = \varphi_n \end{array}\right\} \tag{523}$$

in which L_{ik} are the mesh and mutual inductances, $\dot{q}_k = i_k$ is written as an abbreviation for the time derivative of the mesh charge, and φ_s is the total flux linkage with respect to mesh s due to all the self and mutual inductances on its contour.

If in the system (523) we multiply the first equation by the mesh current $\dot{q}_1$, the second by $\dot{q}_2, \cdots$ and finally the last by $\dot{q}_n$, we obtain

$$\left.\begin{array}{l} L_{11}\dot{q}_1^2 + L_{12}\dot{q}_1\dot{q}_2 + \cdots + L_{1n}\dot{q}_1\dot{q}_n = \varphi_1\dot{q}_1 \\ L_{21}\dot{q}_2\dot{q}_1 + L_{22}\dot{q}_2^2 + \cdots + L_{2n}\dot{q}_2\dot{q}_n = \varphi_2\dot{q}_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ L_{n1}\dot{q}_n\dot{q}_1 + L_{n2}\dot{q}_n\dot{q}_2 + \cdots + L_{nn}\dot{q}_n^2 = \varphi_n\dot{q}_n \end{array}\right\} \tag{523a}$$

The members on the right-hand side, which have the form $\varphi_i\dot{q}_i$, are *twice* the magnetic (kinetic) energies stored in the flux fields of meshes 1 to n respectively at any instant of time. Their sum, therefore, represents twice the total magnetic energy of the system at any instant. The total magnetic energy we shall denote by T, and hence we have

$$T = \tfrac{1}{2}(\varphi_1\dot{q}_1 + \varphi_2\dot{q}_2 + \cdots + \varphi_n\dot{q}_n). \tag{524}$$

If we wish to write this in terms of the inductances L_{ik} directly, we may do so by making use of a double sum. The sum of (523*a*) thus gives

$$T = \tfrac{1}{2}\sum_{\substack{i=1\\k=1}}^{n} L_{ik}\dot{q}_i\dot{q}_k, \tag{524a}$$

in which the summation on k from 1 to n gives the ith row in (523*a*). The expression (524*a*), in which each term involves the square of current, is referred to in mathematics as a **quadratic form.** The matrix of its coefficients

$$\left\|\begin{array}{cccc} L_{11} & L_{12} & \cdots & L_{1n} \\ L_{21} & L_{22} & \cdots & L_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ L_{n1} & L_{n2} & \cdots & L_{nn} \end{array}\right\| \tag{525}$$

is called the **matrix of the quadratic form,** and its determinant the **determinant or discriminant of the quadratic form.**

In a similar manner the total instantaneous potential energy of the network may be set up. In the following equations

$$\left.\begin{array}{l} S_{11}q_1 + S_{12}q_2 + \cdots + S_{1n}q_n = e_{c1} \\ S_{21}q_1 + S_{22}q_2 + \cdots + S_{2n}q_n = e_{c2} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ S_{n1}q_1 + S_{n2}q_2 + \cdots + S_{nn}q_n = e_{cn} \end{array}\right\}, \tag{526}$$

the terms $e_{c1} \ldots e_{cn}$ are the capacitance voltages induced on the contours of the respective meshes. Multiplying these equations respectively by $q_1, q_2, \ldots q_n$ and adding we get *twice* the total instantaneous electrostatic or potential energy of the system. Denoting this energy function by V we have

$$V = \tfrac{1}{2}(e_{c1}q_1 + e_{c2}q_2 + \cdots + e_{cn}q_n), \tag{527}$$

or the quadratic form

$$V = \tfrac{1}{2}\sum_{\substack{i=1\\k=1}}^{n} S_{ik}q_iq_k. \tag{527a}$$

Finally the loss energy may be put into such a form. Regarding the following equations for the induced voltages due to resistances on the various mesh contours

$$\left.\begin{array}{l} R_{11}\dot{q}_1 + R_{12}\dot{q}_2 + \cdots + R_{1n}\dot{q}_n = e_{r1} \\ R_{21}\dot{q}_1 + R_{22}\dot{q}_2 + \cdots + R_{2n}\dot{q}_n = e_{r2} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ R_{n1}\dot{q}_1 + R_{n2}\dot{q}_2 + \cdots + R_{nn}\dot{q}_n = e_{rn} \end{array}\right\}, \tag{528}$$

we note that the first multiplied by $\dot{q}_1$ represents the instantaneous rate of energy loss in mesh 1, the second multiplied by $\dot{q}_2$ represents that for mesh 2, etc. Denoting *half* the total instantaneous rate of energy loss by F, we have

$$F = \tfrac{1}{2}(e_{r1}\dot{q}_1 + e_{r2}\dot{q}_2 + \cdots + e_{rn}\dot{q}_n), \tag{529}$$

or

$$F = \tfrac{1}{2}\sum_{\substack{i=1\\k=1}}^{n} R_{ik}\dot{q}_i\dot{q}_k. \tag{529a}$$

The quadratic forms (524*a*), (527*a*), and (529*a*) have one outstanding characteristic in common, namely, that *they are positive for all values of the variables.* This is quite obviously necessary from physical considerations. It is well known that a quadratic form is positive in the above sense if its determinant and all its principal minors are positive.[1]

[1] See "An Elementary Treatise on the Differential Calculus" by Benjamin Williamson, Longmans Green & Co., 1893, pp. 447-451.

Hence we see that arbitrarily given matrices of L, S, and R coefficients can characterize a physical network *only if their determinants and principal minors are positive.* This necessary condition for physical realizability is also sufficient.[1] Thus for physical networks we always must have

$$\left.\begin{array}{l} \begin{vmatrix} L_{11} & \cdots & L_{1k} \\ \cdots & \cdots & \cdots \\ L_{k1} & \cdots & L_{kk} \end{vmatrix} \geq 0 \\ \begin{vmatrix} R_{11} & \cdots & R_{1k} \\ \cdots & \cdots & \cdots \\ R_{k1} & \cdots & R_{kk} \end{vmatrix} \geq 0 \\ \begin{vmatrix} S_{11} & \cdots & S_{1k} \\ \cdots & \cdots & \cdots \\ S_{k1} & \cdots & S_{kk} \end{vmatrix} \geq 0 \end{array}\right\}; \ (k = 1, 2, \cdots n). \tag{530}$$

In order to correlate the energy forms with the *emf* equations of the network we note that

$$\left.\begin{aligned} \varphi_k &= \frac{\partial T}{\partial \dot{q}_k} \\ e_{rk} &= \frac{\partial F}{\partial \dot{q}_k} \\ e_{ck} &= \frac{\partial V}{\partial q_k} \end{aligned}\right\}. \tag{531}$$

This may be seen for the φ_k's, for example, by differentiating the sum of the left-hand sides of (523a) partially with respect to, say, $\dot{q}_1$. The terms in the first row and column will then add up to give $2\varphi_1$ according to the first equation (523), while the remaining terms will all be zero because they do not contain $\dot{q}_1$. The same treatment applies to the other quadratic forms.

Now since $\dfrac{d\varphi_k}{dt}$ is the inductive countervoltage on the contour of mesh k, the Kirchhoff emf equations for that mesh are

$$\frac{d\varphi_k}{dt} + e_{rk} + e_{ck} = e_k \tag{532}$$

where e_k is the impressed voltage. Using (531) in (532) we see that the

[1] In the case of resistance and elastance parameters this assumes that ideal transformers may be employed. Similarly, mutual inductances are considered "realizable" even with unity coupling coefficients.

dynamic equilibrium conditions for the network in terms of the total instantaneous magnetic, potential, and loss energy functions[1] are

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right) + \frac{\partial F}{\partial \dot{q}_k} + \frac{\partial V}{\partial q_k} = e_k; \ (k = 1, 2, \cdots n). \tag{532a}$$

These are the so-called **Lagrangian equations of motion.**

3. The driving-point impedance function and its properties in terms of the energy functions.[2] Several useful properties regarding driving-point impedances may be derived in terms of the energy functions developed in the previous section. In this connection we are interested, of course, in the steady-state behavior of the network when it is driven from a single point by means of a sinusoidal voltage. In such a case the mesh charges and currents are simple harmonic functions of time. We will write them in the form

$$\left.\begin{aligned} q_k &= \frac{1}{2j\omega}(I_k e^{j\omega t} - \bar{I}_k e^{-j\omega t}) \\ \text{and} \quad \dot{q}_k &= \tfrac{1}{2}(I_k e^{j\omega t} + \bar{I}_k e^{-j\omega t}) \end{aligned}\right\}, \tag{533}$$

where I_k is the complex mesh current amplitude, and the bar denotes the conjugate value.

Our first objective is to form expressions for the corresponding magnetic, potential, and loss energy functions averaged over a period. These are indicated by means of (524a), (527a), and (529a) in the following expressions

$$T_{av} = \frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}} T\,dt = \tfrac{1}{2}\sum_{i,k=1}^{n} L_{ik}\left(\frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}} \dot{q}_i\dot{q}_k\,dt\right),$$

$$V_{av} = \frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}} V\,dt = \tfrac{1}{2}\sum_{i,k=1}^{n} S_{ik}\left(\frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}} q_i q_k\,dt\right),$$

and

$$F_{av} = \frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}} F\,dt = \tfrac{1}{2}\sum_{i,k=1}^{n} R_{ik}\left(\frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}} \dot{q}_i\dot{q}_k\,dt\right).$$

Substituting from (533) we get, for example, for the first of these

$$T_{av} = \frac{1}{8}\sum_{i,k=1}^{n} L_{ik}\cdot\frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}} (I_i e^{j\omega t} + \bar{I}_i e^{-j\omega t})(I_k e^{j\omega t} + \bar{I}_k e^{-j\omega t})\,dt.$$

[1] Here it should be remembered that the loss function represents half the instantaneous loss, while the potential and kinetic energy functions represent the total energies.

[2] The method used in this section is due to H. W. Bode.

Evaluation of this expression gives

$$T_{\text{av}} = \frac{1}{8} \sum_{i,k=1}^{n} L_{ik}(I_i\bar{I}_k + \bar{I}_i I_k).$$

This may be contracted further if we make use of the fact that $L_{ik} = L_{ki}$ and note that since i and k are merely summation indices we have

$$\sum_{i,k=1}^{n} L_{ik}\bar{I}_i I_k = \sum_{i,k=1}^{n} L_{ki} I_k \bar{I}_i = \sum_{i,k=1}^{n} L_{ik} I_i \bar{I}_k.$$

Thus we finally get

$$T_{\text{av}} = \frac{1}{4} \sum_{i,k=1}^{n} L_{ik} I_i \bar{I}_k, \tag{524b}$$

and similarly for the average potential and loss energies

$$V_{\text{av}} = \frac{1}{4\omega^2} \sum_{i,k=1}^{n} S_{ik} I_i \bar{I}_k, \tag{527b}$$

and

$$F_{\text{av}} = \frac{1}{4} \sum_{i,k=1}^{n} R_{ik} I_i \bar{I}_k. \tag{529b}$$

Suppose that the network is excited from its first mesh by means of a generator whose voltage amplitude is taken as a reference and denoted by E_1. The equilibrium conditions are then

$$\left.\begin{array}{l} b_{11}I_1 + b_{12}I_2 + \cdots + b_{1n}I_n = E_1 \\ b_{21}I_1 + b_{22}I_2 + \cdots + b_{2n}I_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ b_{n1}I_1 + b_{n2}I_2 + \cdots + b_{nn}I_n = 0 \end{array}\right\}. \tag{534}$$

where

$$b_{ik} = R_{ik} + j\omega\left(L_{ik} - \frac{S_{ik}}{\omega^2}\right). \tag{534a}$$

If we multiply the first of these equations by $\bar{I}_1$, the second by $\bar{I}_2$, etc., and finally the last by $\bar{I}_n$ and add, we find

$$E_1\bar{I}_1 = 4\{F_{\text{av}} + j\omega(T_{\text{av}} - V_{\text{av}})\}.$$

Since E_1 has been chosen as a reference, and the energy functions are real, this gives

$$I_1 = \frac{4\{F_{\text{av}} - j\omega(T_{\text{av}} - V_{\text{av}})\}}{E_1}.$$

This divided by E_1 is the driving-point admittance function for the network, thus

$$Y_{11} = \frac{4\{F_{\text{av}} - j\omega(T_{\text{av}} - V_{\text{av}})\}}{E_1^2}. \tag{535}$$

The driving-point impedance function is, of course, the reciprocal of this expression.

Since the real part of an admittance is an even function of ω and its imaginary part odd (see footnote, page 214), it follows that the average energies are all even functions of ω. The form (535) is used to deduce a number of useful theorems regarding driving-point functions. Here the positiveness of the energy functions forms the guiding principle. The positiveness of the slope of a driving-point reactance function (which was utilized in the preceding chapter) may thus be proved. Instead of using the form (535) directly for this purpose, it is more convenient to prove this property in another way.

Consider the input voltage E_1 in (534) constant and differentiate the equations with respect to ω. If we indicate such differentiation by means of primes, we get

$$\left.\begin{array}{l} b_{11}I_1' + b_{12}I_2' + \cdots + b_{1n}I_n' = -H_1 \\ b_{21}I_1' + b_{22}I_2' + \cdots + b_{2n}I_n' = -H_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ b_{n1}I_1' + b_{n2}I_2' + \cdots + b_{nn}I_n' = -H_n \end{array}\right\}, \tag{536}$$

where

$$H_\nu = b_{\nu 1}'I_1 + b_{\nu 2}'I_2 + \cdots + b_{\nu n}'I_n. \tag{536a}$$

Solving (536) for I_1' we have

$$I_1' = -\sum_{\nu=1}^{n} \frac{B_{\nu 1}H_\nu}{D} = -\sum_{\nu=1}^{n} \frac{B_{1\nu}H_\nu}{D}.$$

But by (534)

$$\frac{B_{1\nu}}{D} = \frac{I_\nu}{E_1},$$

so that

$$I_1' = -\frac{1}{E_1}\sum_{\nu=1}^{n} H_\nu I_\nu = \frac{1}{E_1}\sum_{\nu=1}^{n} H_\nu \bar{I}_\nu$$

because in the non-dissipative case $\bar{I}_\nu = -I_\nu$ when E_1 is taken as reference. Thus we have

$$E_1 I_1' = \sum_{\nu=1}^{n} H_\nu \bar{I}_\nu.$$

Denoting the driving-point reactance by jX_{11}, we can write

$$I_1 = \frac{E_1}{jX_{11}},$$

and get

$$I_1' = \frac{jE_1}{X_{11}^2}\left(\frac{dX_{11}}{d\omega}\right),$$

so that

$$E_1 I_1' = j\left(\frac{E_1}{X_{11}}\right)^2 \cdot \left(\frac{dX_{11}}{d\omega}\right) = \sum_{\nu=1}^{n} H_\nu \bar{I}_\nu.$$

Substituting (536*a*) into the right-hand side of this equation and noting that for the reactive case

$$b_{ik}' = j\left(L_{ik} + \frac{S_{ik}}{\omega^2}\right),$$

the use of (524*b*) and (527*b*) gives

$$\frac{dX_{11}}{d\omega} = 4\left(\frac{X_{11}}{E_1}\right)^2 \cdot (T_{av} + V_{av}). \tag{537}$$

Utilizing the result (535) for zero dissipation, this may be put into the form

$$\frac{dX_{11}}{d\omega} = \frac{E_1^2}{4\omega^2} \cdot \frac{T_{av} + V_{av}}{(T_{av} - V_{av})^2}. \tag{537a}$$

Since the average energy functions are positive, *the slope of the reactance function must be positive and greater than zero.*[1]

4. Transformations which keep an impedance invariant. In the light of the above ideas we may approach the problem of finding equivalent networks by means of linear transformations. The general method was first suggested by Cauer[2] and later carried through by Howitt,[3] who showed that the parameters of the transformed network were obtained from those of the original by applying the desired transformations to the quadratic forms.

If the mesh charges of the given network are denoted by q_k and those of the transformed network by q_k', then the following linear transformation

$$\left.\begin{aligned} q_1 &= q_1' \\ q_2 &= \alpha_{21}q_1' + \alpha_{22}q_2' + \cdots + \alpha_{2n}q_n' \\ &\cdots\cdots\cdots\cdots \\ q_n &= \alpha_{n1}q_1' + \alpha_{n2}q_2' + \cdots + \alpha_{nn}q_n' \end{aligned}\right\} \tag{538}$$

[1] By the duality principle (discussed in section 9 of this chapter) it follows that a *susceptance* function has the same property. The separation property of the zeros and poles of these functions (utilized in the preceding chapter) is a direct consequence of their positive slope, and since the latter must be greater than zero, the zeros of both functions and hence the zeros and poles of either must be simple, as is clear from the consideration that at a zero of higher multiplicity the function must have zero slope.

[2] W. Cauer, "Vierpole," E. N. T., Band 6, Heft 7, 1929, p. 272. A further development is given in a later paper entitled, "Ideale Transformatoren und lineare Transformationen," E. N. T., Band 9, Heft 5, 1932, p. 157.

[3] N. Howitt, "Group Theory and the Electric Circuit," Phys. Rev., Vol. 37, Ser. 2, No. 12, 1931, p. 1583.

will leave the charge and hence the current in mesh 1 invariant. If this mesh is the driving point, then the network which is obtained from the original by means of the transformation (538) will have the same driving-point impedance. The only restriction on the coefficients α_{sr} of the transformation (538) is that they shall be real.

This transformation can be more conveniently written in the form

$$q_i = \sum_{r=1}^{n} \alpha_{ir} q_r'; \ (i = 1, 2, \cdots n), \tag{538a}$$

with the stipulation that

$$\alpha_{11} = 1; \ \alpha_{12} = \alpha_{13} = \cdots = \alpha_{1n} = 0. \tag{539}$$

In order to determine the parameters of the transformed network, the transformation (538a) may be substituted into the quadratic forms giving the kinetic, potential, and loss functions for the original network. Suppose these are given by (524a), (527a), and (529a) respectively. For q_i we substitute (538a). For q_k we use

$$q_k = \sum_{s=1}^{n} \alpha_{ks} q_s'; \ (k = 1, 2, \cdots n) \tag{540}$$

so as not to confuse indices. Then we get for the new network

$$\left.\begin{aligned} T' &= \tfrac{1}{2}\sum_{i,k=1}^{n} \sum_{r,s=1}^{n} \alpha_{ir}\alpha_{ks} L_{ik} \dot{q}_r' \dot{q}_s' \\ V' &= \tfrac{1}{2}\sum_{i,k=1}^{n} \sum_{r,s=1}^{n} \alpha_{ir}\alpha_{ks} S_{ik} q_r' q_s' \\ F' &= \tfrac{1}{2}\sum_{i,k=1}^{n} \sum_{r,s=1}^{n} \alpha_{ir}\alpha_{ks} R_{ik} \dot{q}_r' \dot{q}_s' \end{aligned}\right\}. \tag{541}$$

Inverting the order of summation these are

$$\left.\begin{aligned} T' &= \tfrac{1}{2}\sum_{r,s=1}^{n} \left(\sum_{i,k=1}^{n} \alpha_{ir}\alpha_{ks} L_{ik} \right) \dot{q}_r' \dot{q}_s' \\ V' &= \tfrac{1}{2}\sum_{r,s=1}^{n} \left(\sum_{i,k=1}^{n} \alpha_{ir}\alpha_{ks} S_{ik} \right) q_r' q_s' \\ F' &= \tfrac{1}{2}\sum_{r,s=1}^{n} \left(\sum_{i,k=1}^{n} \alpha_{ir}\alpha_{ks} R_{ik} \right) \dot{q}_r' \dot{q}_s' \end{aligned}\right\}. \tag{541a}$$

The parameters of the transformed network are those quantities enclosed in parentheses. Denoting the new inductance, elastance, and resistance parameters by L_{rs}', S_{rs}', and R_{rs}', respectively, we have

$$\left.\begin{aligned} L_{rs}' &= \sum_{i,k=1}^{n} \alpha_{ir}\alpha_{ks} L_{ik} \\ S_{rs}' &= \sum_{i,k=1}^{n} \alpha_{ir}\alpha_{ks} S_{ik} \\ R_{rs}' &= \sum_{i,k=1}^{n} \alpha_{ir}\alpha_{ks} R_{ik} \end{aligned}\right\}. \tag{541b}$$

Recalling the methods of matrix algebra discussed in Chapter IV, this result can be obtained in another way. For example, in matrix form the equations (523) are

$$\left\| \begin{matrix} L_{11} & \cdots & L_{1n} \\ \cdot & \cdots & \cdot \\ L_{n1} & \cdots & L_{nn} \end{matrix} \right\| \times \left\| \begin{matrix} \dot{q}_1 \\ \vdots \\ \dot{q}_n \end{matrix} \right\| = \left\| \begin{matrix} \varphi_1 \\ \vdots \\ \varphi_n \end{matrix} \right\|, \tag{523b}$$

and since

$$\| \dot{q}_1 \cdots \dot{q}_n \| \times \left\| \begin{matrix} \varphi_1 \\ \vdots \\ \varphi_n \end{matrix} \right\| = \varphi_1\dot{q}_1 + \cdots + \varphi_n\dot{q}_n,$$

we have

$$T = \| \dot{q}_1 \cdots \dot{q}_n \| \times \left\| \begin{matrix} L_{11} & \cdots & L_{1n} \\ \cdot & \cdots & \cdot \\ L_{n1} & \cdots & L_{nn} \end{matrix} \right\| \times \left\| \begin{matrix} \dot{q}_1 \\ \vdots \\ \dot{q}_n \end{matrix} \right\|. \tag{524c}$$

Writing the time derivative of the transformation (538a) in matrix form

$$\left\| \begin{matrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \cdot & \cdots & \cdot \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{matrix} \right\| \times \left\| \begin{matrix} \dot{q}_1' \\ \vdots \\ \dot{q}_n' \end{matrix} \right\| = \left\| \begin{matrix} \dot{q}_1 \\ \vdots \\ \dot{q}_n \end{matrix} \right\|, \tag{538b}$$

we note that reversing the order of multiplication and interchanging rows and columns gives

$$\| \dot{q}_1' \cdots \dot{q}_n' \| \times \left\| \begin{matrix} \alpha_{11} & \cdots & \alpha_{n1} \\ \cdot & \cdots & \cdot \\ \alpha_{1n} & \cdots & \alpha_{nn} \end{matrix} \right\| = \| \dot{q}_1 \cdots \dot{q}_n \|. \tag{538c}$$

Substituting the last two equations into (524c) gives

$$T' = \| \dot{q}_1' \cdots \dot{q}_n' \| \times \left\| \begin{matrix} L_{11}' & \cdots & L_{1n}' \\ \cdot & \cdots & \cdot \\ L_{n1}' & \cdots & L_{nn}' \end{matrix} \right\| \times \left\| \begin{matrix} \dot{q}_1' \\ \vdots \\ \dot{q}_n' \end{matrix} \right\|, \tag{524d}$$

where

$$\left\| \begin{matrix} L_{11}' & \cdots & L_{1n}' \\ \cdot & \cdots & \cdot \\ L_{n1}' & \cdots & L_{nn}' \end{matrix} \right\| = \left\| \begin{matrix} \alpha_{11} & \cdots & \alpha_{n1} \\ \cdot & \cdots & \cdot \\ \alpha_{1n} & \cdots & \alpha_{nn} \end{matrix} \right\| \times \left\| \begin{matrix} L_{11} & \cdots & L_{1n} \\ \cdot & \cdots & \cdot \\ L_{n1} & \cdots & L_{nn} \end{matrix} \right\| \times \left\| \begin{matrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \cdot & \cdots & \cdot \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{matrix} \right\|. \tag{542}$$

This last equation gives the new inductance matrix in terms of the old and the transformation matrix. Note that the first matrix on the right-hand side of (542) is the transformation matrix *with its rows written as columns.* This is called the **conjugate matrix.** Denoting the transformation matrix by A, its conjugate by A', the original inductance matrix by L, and the new inductance matrix by L', the result

(542) and the similar results for elastance and resistance may be written

$$\left.\begin{array}{l} \mathsf{L}' = \mathsf{A}'\,\mathsf{L}\,\mathsf{A} \\ \mathsf{S}' = \mathsf{A}'\,\mathsf{S}\,\mathsf{A} \\ \mathsf{R}' = \mathsf{A}'\,\mathsf{R}\,\mathsf{A} \end{array}\right\}, \tag{543}$$

where S and R are the elastance and resistance matrices

$$\mathsf{S} = \begin{Vmatrix} S_{11} & \cdots & S_{1n} \\ \cdots & \cdots & \cdots \\ S_{n1} & \cdots & S_{nn} \end{Vmatrix}; \quad \mathsf{R} = \begin{Vmatrix} R_{11} & \cdots & R_{1n} \\ \cdots & \cdots & \cdots \\ R_{n1} & \cdots & R_{nn} \end{Vmatrix}, \tag{544}$$

and the primes refer to the new network.

These matrix equations are identical with the results (541*b*). This the student should verify by writing out the above indicated operations more fully.

Regarding the physical realizability of the transformed network, we note that if $|\alpha_{ik}|$ denotes the determinant of the transformation, $|L_{ik}|$ that of the inductance matrix, and $|L_{ik}'|$ that of the transformed inductance matrix, we have

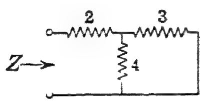


FIG. 84.—Driving-point resistance network.

$$|L_{ik}'| = |\alpha_{ik}|^2 \cdot |L_{ik}|, \tag{545}$$

and similar relations with respect to elastance and resistance. Since the transformation is real, $|\alpha_{ik}|^2$ is positive, and hence the transformed network has quadratic forms with positive determinants if the original ones are positive. The quadratic forms of the transformed network are positive definite when obtained from positive definite forms by means of a real non-singular transformation.[1] Hence if we begin with a physical network we should always arrive at realizable networks by means of real non-singular transformations.

This point needs a few additional remarks, however. Although a given matrix has a positive determinant with positive principal minors, the corresponding network may not be realizable without the aid of ideal transformers. As a simple illustration consider the resistance network of Fig. 84. This has the resistance matrix

$$\mathsf{R} = \begin{Vmatrix} 6 & -4 \\ -4 & 7 \end{Vmatrix}.$$

[1] This follows from the fact that a real non-singular transformation preserves the rank and the signature of a quadratic form. See M. Bôcher, Theorem 2, p. 148.

Suppose we transform this by means of the matrix

$$\mathsf{A} = \begin{Vmatrix} 1 & 0 \\ 2 & -3 \end{Vmatrix},$$

so that

$$\mathsf{R}' = \begin{Vmatrix} 1 & 2 \\ 0 & -3 \end{Vmatrix} \times \begin{Vmatrix} 6 & -4 \\ -4 & 7 \end{Vmatrix} \times \begin{Vmatrix} 1 & 0 \\ 2 & -3 \end{Vmatrix} = \begin{Vmatrix} 18 & -30 \\ -30 & 63 \end{Vmatrix}.$$

The determinant of this matrix is positive, but the total resistance in the first mesh is less than that common to the two meshes. This requires negative resistance unless an ideal transformer is made use of. For example, if the network is preceded by an ideal transformer having a ratio $a : 1$, then the resistance matrix may be written

$$\mathsf{R}' = \begin{Vmatrix} a^2R_{11} & aR_{12} \\ aR_{21} & R_{22} \end{Vmatrix} = \begin{Vmatrix} 18 & -30 \\ -30 & 63 \end{Vmatrix}.$$

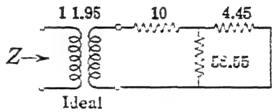


FIG. 85.—Driving-point resistance network equivalent to that in Fig. 84.

Thus we have

$$a^2R_{11} = 18; \; aR_{12} = -30; \; R_{22} = 63,$$

and if we wish to avoid negative resistance in the first mesh we must have

$$R_{11} + R_{12} \geq 0.$$

Here any positive value may be assumed. Suppose we let

$$R_{11} + R_{12} = 10; \; R_{12} = 10 - R_{11}.$$

Then

$$a(10 - R_{11}) = -30; \; a^2R_{11} = 18,$$

which gives

$$R_{11} = 68.55; \; a = \frac{1}{1.95}.$$

The network is that shown in Fig. 85, which contains no negative elements.

Such methods can always be carried out in order to get rid of negative elements. In more extensive networks, this process may introduce an undue number of ideal transformers. Practically such networks are of little if any value. At all events we should attempt to avoid ideal transformers wherever possible. Transformations which lead to negative elements unless ideal transformers are used are impracticable.

It is generally difficult to tell by inspection of a given transformation

matrix whether it will lead to negative elements when ideal transformers are omitted. These negative elements, it should be remembered, will occur whenever the total inductance, elastance, or resistance on a mesh contour is less than the sum in the various mutual branches of that mesh. It would seem that, if a given transformation were broken up into a succession of simpler ones, then it might be easier to see where each step led to, and thus undesirable results could be avoided. Such a step-by-step procedure would also be more useful if we were trying to arrive at a definite form as a result. How this may be done will now be discussed.

5. Transformation by matrix manipulation. The idea of this method is to change the given matrices successively by means of operations each of which preserves the invariance of the desired impedance function.

In order to formulate these operations, let us consider a transformation with the matrix

$$\mathsf{A} = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix}, \tag{546}$$

which is equivalent to

$$\left.\begin{aligned} q_1 &= q_1' \\ q_2 &= q_2' \\ \cdot\ \cdot & \ \cdot\ \cdot\ \cdot \\ q_n &= q_n' \end{aligned}\right\}. \tag{546a}$$

This is called the **identity transformation.** If it is applied to the system of equations (523*a*), for example, it will merely replace all the q_k's by q_k''s.

Suppose now that we multiply the kth column in A by a factor m. This is the same as having

$$q_k = mq_k', \tag{547}$$

the rest of the equations (546*a*) remaining intact. The effect which this will have on say the L-matrix is seen from (523*a*) by replacing q_k by mq_k' and placing primes on the rest of the q's. Since q_k is a factor of the kth *row*, and again of the kth *column*, we see that multiplying a column in A by a factor m has the effect of multiplying the corresponding *row and column* in L by this factor. The same applies to the R and S matrices.

We can, therefore, successively transform a given set of matrices by repeatedly multiplying rows and columns in them by any factors. So

long as we avoid the sth *row* and *column*, the charge q_s remains invariant, and hence the impedance with respect to that mesh is unaffected. This assumes, of course, that the driving force is located in the sth mesh, otherwise no driving-point impedance would be represented.

Such operations alone, however, are rather limited in the results which they can produce. A wider range is obtained if we recognize that a subsequent substitution

$$\left.\begin{array}{l} q_1' = q_1'' \\ \cdots\cdots \\ q_k' = q_k'' + q_s'' \\ \cdots\cdots\cdots \\ q_n' = q_n'' \end{array}\right\} \tag{548}$$

is equivalent to adding the kth column in the transformation matrix to the sth column, and its effect upon the L, R, and S matrices is to add the kth column to the sth column, and then add the kth row to the sth row. In fact, the combination of multiplying by factors and adding rows and columns, if repeatedly applied, will produce any desired transformation. It should be noted that if q_r is to remain invariant, then the sth row may be added to the rth row and the sth column to the rth column, *but not the reverse*. This is clear from (548). It is also useful to note that, by a combination of adding and subtracting, any two columns in the transformation matrix may be interchanged. Hence interchanging of rows and columns in the L, R, and S matrices is a legitimate procedure so long as the rth row and column are not involved where q_r is the mesh charge which is to remain invariant.

6. Extension to more pairs of terminals. The above transformation (538) or its equivalent (538a) with the stipulation (539) leads to other networks for which the charge and current in mesh 1 are the same. It is to be supposed that the network is driven by a generator located in mesh 1. The driving-point impedance presented at the generator terminals is thus invariant to such a transformation. We may easily extend the linear transformation method to get equivalents to networks with more than one pair of terminals. For example, if we have a four-terminal network with the two pairs of terminals $1 - 1'$ and $2 - 2'$ located in meshes 1 and 2 respectively, then a transformation which will leave the currents in both of these invariant will evidently leave the external behavior of the network between its pairs of terminals invariant. Such a transformation is also specified by (538a), but with the stipulation that

$$\left.\begin{array}{r} \alpha_{11} = 1;\ \alpha_{12} = \alpha_{13} = \cdots = \alpha_{1n} = 0 \\ \alpha_{22} = 1;\ \alpha_{21} = \alpha_{23} = \cdots = \alpha_{2n} = 0 \end{array}\right\}. \tag{549}$$

In a similar manner the method can be extended to apply to networks with any number of pairs of terminals. Obviously the number of meshes must exceed the number of pairs of terminals, otherwise the transformation will be the identity transformation (546), and hence no equivalents can be found by this method.

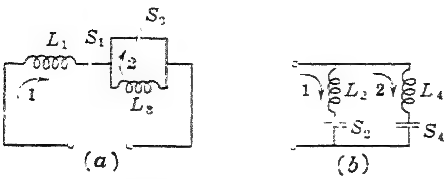


FIG. 86.—Potentially equivalent reactive driving-point networks according to Foster.

7. Illustrations. As an example for the two-terminal case, consider the reactive networks of Fig. 86. These we recognize as potentially equivalent. Suppose the parameters of the network (a) are

$$\left.\begin{array}{l} L_1 = 1;\ S_1 = 1 \\ L_3 = 3;\ S_3 = 3 \end{array}\right\},$$

and it is desired to find the parameters of (b) which will make its driving-point impedance equal to that of (a). For the latter we have by inspection

$$\mathsf{L} = \begin{Vmatrix} 4 & -3 \\ -3 & 3 \end{Vmatrix}; \quad \mathsf{S} = \begin{Vmatrix} 1 & 0 \\ 0 & 3 \end{Vmatrix}.$$

The transformation matrix has the form

$$\mathsf{A} = \begin{Vmatrix} 1 & 0 \\ \alpha_{21} & \alpha_{22} \end{Vmatrix}.$$

Applying the matrix multiplication method (542) we have

$$\mathsf{L}' = \begin{Vmatrix} 1 & \alpha_{21} \\ 0 & \alpha_{22} \end{Vmatrix} \times \begin{Vmatrix} 4 & -3 \\ -3 & 3 \end{Vmatrix} \times \begin{Vmatrix} 1 & 0 \\ \alpha_{21} & \alpha_{22} \end{Vmatrix} = \begin{Vmatrix} 4 - 6\alpha_{21} + 3\alpha_{21}^2 & 3\alpha_{22}(\alpha_{21} - 1) \\ 3\alpha_{22}(\alpha_{21} - 1) & 3\alpha_{22}^2 \end{Vmatrix},$$

and

$$\mathsf{S}' = \begin{Vmatrix} 1 & \alpha_{21} \\ 0 & \alpha_{22} \end{Vmatrix} \times \begin{Vmatrix} 1 & 0 \\ 0 & 3 \end{Vmatrix} \times \begin{Vmatrix} 1 & 0 \\ \alpha_{21} & \alpha_{22} \end{Vmatrix} = \begin{Vmatrix} 1 + 3\alpha_{21}^2 & 3\alpha_{21}\alpha_{22} \\ 3\alpha_{21}\alpha_{22} & 3\alpha_{22}^2 \end{Vmatrix}.$$

Now we note that the network (b) of Fig. 86 has inductance and elastance matrices of the form

$$\mathsf{L} = \begin{Vmatrix} L_2 & -L_2 \\ -L_2 & L_2 + L_4 \end{Vmatrix}; \quad \mathsf{S} = \begin{Vmatrix} S_2 & -S_2 \\ -S_2 & S_2 + S_4 \end{Vmatrix}.$$

These forms are characterized by the fact that the elements in the first rows are the negatives of each other. Hence in the matrices L' and S' we must have

$$4 - 6\alpha_{21} + 3\alpha_{21}^2 = -3\alpha_{22}(\alpha_{21} - 1),$$

and

$$1 + 3\alpha_{21}^2 = -3\alpha_{21}\alpha_{22}.$$

The coefficients α_{21} and α_{22} must be determined so that these relations will hold simultaneously. Eliminating we find

$$\alpha_{21}^2 - \alpha_{21} - \tfrac{1}{3} = 0,$$

from which

$$\alpha_{21} = 1.264 \text{ or } \alpha_{21} = -0.264.$$

Only one of these roots need be considered. The other will give the

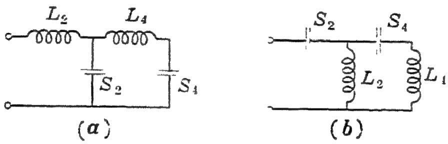


FIG. 87.—Potentially equivalent reactive driving-point networks according to Cauer.

same form of network with the two resonant components interchanged. Using the second root we find

$$\alpha_{22} = 1.528,$$

and hence

$$\mathsf{L}' = \begin{Vmatrix} 5.79 & -5.79 \\ -5.79 & 7 \end{Vmatrix}; \quad \mathsf{S}' = \begin{Vmatrix} 1.21 & -1.21 \\ -1.21 & 7 \end{Vmatrix}.$$

The parameters of the desired network are, therefore,

$$L_2 = 5.79; \; L_4 = 1.21; \; S_2 = 1.21; \; S_4 = 5.79.$$

The networks (a) and (b) of Fig. 87 are also potentially equivalent to those of Fig. 86. For the first of these we have

$$\mathsf{L} = \begin{Vmatrix} L_2 & 0 \\ 0 & L_4 \end{Vmatrix}; \quad \mathsf{S} = \begin{Vmatrix} S_2 & -S_2 \\ -S_2 & S_2 + S_4 \end{Vmatrix}.$$

Comparison with L' and S' above shows that

$$\left.\begin{aligned} 3\alpha_{22}(\alpha_{21} - 1) &= 0 \\ 1 + 3\alpha_{21}^2 &= -3\alpha_{21}\alpha_{22} \end{aligned}\right\}$$

Since α_{22} cannot be zero, it follows that $\alpha_{21} = 1$, and hence $\alpha_{22} = -1.33$. The matrices L' and S', therefore, become

$$\mathsf{L}' = \begin{Vmatrix} 1 & 0 \\ 0 & 5.33 \end{Vmatrix}; \quad \mathsf{S}' = \begin{Vmatrix} 4 & -4 \\ -4 & 5.33 \end{Vmatrix},$$

and comparison with L and S for this network shows that

$$L_2 = 1;\ L_4 = 5.33;\ S_2 = 4;\ S_4 = 1.33.$$

Finally for network (b) of Fig. 87 we have

$$\mathsf{L} = \begin{Vmatrix} L_2 & -L_2 \\ -L_2 & L_2 + L_4 \end{Vmatrix}; \quad \mathsf{S} = \begin{Vmatrix} S_2 & 0 \\ 0 & S_4 \end{Vmatrix}.$$

In order that L' and S' may have these forms it is necessary that

$$4 - 6\alpha_{21} + 3\alpha_{21}^2 = -3\alpha_{22}(\alpha_{21} - 1)$$
$$3\alpha_{21}\alpha_{22} = 0,$$

from which we have

$$\alpha_{21} = 0;\ \alpha_{22} = -1.33.$$

Hence

$$\mathsf{L}' = \begin{Vmatrix} 4 & -4 \\ -4 & 5.33 \end{Vmatrix}; \quad \mathsf{S}' = \begin{Vmatrix} 1 & 0 \\ 0 & 5.33 \end{Vmatrix},$$

so that for network (b)

$$L_2 = 4;\ L_4 = 1.33;\ S_2 = 1;\ S_4 = 5.33.$$

The four canonic forms for this simple driving-point reactance are, therefore, easily found. In fact, the method of linear transformation is simpler here than Foster's and Cauer's methods. In more extensive cases, more coefficients α_{ik} will have to be determined simultaneously. The process may then become equally as long as that discussed in the previous chapter. However, whereas the Foster and Cauer methods are applicable only to networks involving two kinds of elements, the present method of determining network equivalents is not so limited.

8. Transformation to normal coordinates. In Chapter VII of Volume I we discussed the transient solution for the general network and showed that there exist certain fictitious normal meshes or coordinates in which the natural modes of oscillation and decay are isolated. Since these normal coordinates have arbitrary directions, they in general have projections upon all the actual coordinates or meshes, so that the natural oscillations in these are a linear superposition of all the normal

modes. This fact involves all the integration constants in each mesh current expression so that their evaluation becomes a process of solving a system of simultaneous equations.

If it were possible to transform at the outset from the actual to the normal coordinates, then the evaluation of integration constants would reduce to a process of separately finding the amplitude of oscillation in each normal coordinate from the corresponding initial charge and current. These separate determinations would involve the same procedure as is applied for the determination of the integration constants in the single mesh case.

Such a process of transformation, which has much in common with the matters discussed above, seems very promising from the standpoint of simplifying network solutions. Actually, the difficulty of finding the proper transformation matrix involves about as much labor as is afterward saved by its use, so that little if anything is gained. Added to this is the fact that the transformation is impossible except in certain special instances. These are of sufficient interest, however, to warrant a brief discussion here.

Equation (568) on page 266 of Volume I gives the general expression for the direction cosines of a normal coordinate for the mode p_ν. The quantities $C_{ik}^{(\nu)}$ are the minors of the corresponding coefficients (357) which are the elements of the determinant (360) discussed in Volume I on pages 174 and 175 in connection with the homogeneous condition equations (359). These are written in terms of mesh currents. In order to parallel the work of the present chapter, we wish to change over to the mesh charge basis. This is easily done by multiplying the coefficients of the modular determinant by p.

Considering first the general non-dissipative network, the coefficients of the modular determinant for the mesh charge basis are

$$k_{ik} = L_{ik}p^2 + S_{ik}, \tag{550}$$

and the determinantal equation is

$$D(p) = \begin{vmatrix} k_{11} & \cdots & k_{1n} \\ \cdot\ \cdot\ \cdot & & \cdot \\ k_{n1} & \cdots & k_{nn} \end{vmatrix} = 0. \tag{551}$$

This is a polynomial involving only even powers of p, and is of the $2n$th degree. Considering p^2 as the variable, the roots are all negative real numbers. We shall write these as

so that
$$\left.\begin{aligned} p_\nu^2 &= -g_\nu^2 \\ p_\nu &= \pm jg_\nu \end{aligned}\right\} \quad (\nu = 1,2, \cdots n) \tag{552}$$

are all in the form of pairs of conjugate imaginaries. Each pair determines *one* set of $k_{ik}^{(\nu)}$ according to (550), and hence also one set of $K_{ik}^{(\nu)}$, where the latter are the corresponding minors of the determinant (551).

The direction cosines of the normal coordinates are then

$$l_{k\nu} = \frac{K_{ik}^{(\nu)}}{\sqrt{(K_{i1}^{(\nu)})^2 + (K_{i2}^{(\nu)})^2 + \cdots + (K_{in}^{(\nu)})^2}}, \tag{553}$$

where i is a fixed but arbitrary integer. These are all purely real values. Hence the complete transient mesh charge solutions have the form

$$q_k = \sum_{\nu=1}^{n} l_{k\nu} q_\nu'; \quad (k = 1,2, \ldots n), \tag{554}$$

where

$$q_\nu' = N_\nu e^{jg_\nu t} + \overline{N}_\nu e^{-jg_\nu t} \tag{555}$$

is the transient charge in the normal coordinate corresponding to the mode p_ν, and $2|N_\nu|$ is the amplitude of this oscillation. The relation (554) is, therefore, the desired transformation. The matrix.

$$\mathsf{A} = \begin{Vmatrix} l_{11} & \cdots & l_{1n} \\ \cdot & & \cdot\ \cdot \\ l_{n1} & \cdots & l_{nn} \end{Vmatrix} \tag{556}$$

with coefficients given by (553), simultaneously transforms the quadratic forms representing the potential and kinetic energies for the general non-dissipative network to their normal forms. This means that if the inductance and elastance matrices of (523) and (526) are transformed by means of (556), the resulting matrices will have elements only on their principal diagonals, thus

$$\begin{Vmatrix} l_{11} \cdots l_{n1} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ l_{1n} \cdots l_{nn} \end{Vmatrix} \times \begin{Vmatrix} L_{11} \cdots L_{1n} \\ \cdot \quad \cdot\ \cdot \\ L_{n1} \cdots L_{nn} \end{Vmatrix} \times \begin{Vmatrix} l_{11} \cdots l_{1n} \\ \cdot \quad \cdot\ \cdot \\ l_{n1} \cdots l_{nn} \end{Vmatrix} = \begin{Vmatrix} L_{11}' & 0 & \cdots & 0 \\ 0 & L_{22}' & 0 \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & L_{nn}' \end{Vmatrix}, \tag{557}$$

$$\begin{Vmatrix} l_{11} \cdots l_{n1} \\ \cdot\ \ \cdot\ \cdot\ \cdot \\ l_{1n} \cdots l_{nn} \end{Vmatrix} \times \begin{Vmatrix} S_{11} \cdots S_{1n} \\ \cdot\ \cdot\ \ \cdot\ \cdot\ \cdot \\ S_{n1} \cdots S_{nn} \end{Vmatrix} \times \begin{Vmatrix} l_{11} \cdots l_{1n} \\ \cdot\ \ \cdot\ \cdot\ \cdot \\ l_{n1} \cdots l_{nn} \end{Vmatrix} = \begin{Vmatrix} S_{11}' & 0 & \cdots & 0 \\ 0 & S_{22}' & 0 \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & S_{nn}' \end{Vmatrix} \tag{558}$$

In these we have

$$S_{\nu\nu}' = g_\nu^2 L_{\nu\nu}', \tag{559}$$

which is clear from the fact that after transformation the determinantal equation has the form

$$\begin{vmatrix} (S_{11}' - g_1^2L_{11}') & 0 & \cdots & 0 \\ 0 & (S_{22}' - g_2^2L_{22}') & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & (S_{nn}' - g_n^2L_{nn}') \end{vmatrix} = 0. \quad (560)$$

The quadratic forms for the potential and kinetic energies are reduced, therefore, to sums of squares

$$T' = \tfrac{1}{2}\sum_{\nu=1}^{n} L_{\nu\nu}'\dot{q}_\nu'^2; \quad V' = \tfrac{1}{2}\sum_{\nu=1}^{n} S_{\nu\nu}'q_\nu'^2. \quad (561)$$

The process of evaluating integration constants is to invert the transformation (554) so that the initial transient charges and currents in the normal coordinates may be calculated from the discrepancies between the actual initial charges and currents and the corresponding initial steady-state values. In terms of the $(q_\nu')_{t=0}$ and $(\dot{q}_\nu')_{t=0}$ so determined, the normal amplitudes N_ν in (555) are separately evaluated. Substitution back into (554) then gives the transient mesh charges, and their derivatives are the corresponding currents.

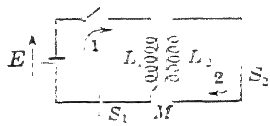


FIG. 88.—Network used as example in text to illustrate the transformation to normal coordinates.

As an illustration consider the network of Fig. 88 with the parameter values

$$L_1 = 10^{-3};\ S_1 = 10^9;\ L_2 = 3\cdot10^{-3};\ S_2 = 1.5\cdot10^9;\ M = 10^{-3}.$$

Here

$$k_{11} = L_1p^2 + S_1;\ k_{12} = Mp^2;\ k_{22} = L_2p^2 + S_2.$$

The determinantal equation is

$$p^4 + 2.25\cdot10^{12}p^2 + 0.75\cdot10^{24} = 0,$$

which has the roots

$$p_1^2 = -1.842\cdot10^{12};\ p_2^2 = -0.408\cdot10^{12}.$$

Hence we have

$$k_{11}^{(1)} = -0.842\cdot10^9;\ k_{12}^{(1)} = -1.842\cdot10^9;\ k_{22}^{(1)} = -4.026\cdot10^9$$
$$k_{11}^{(2)} = 0.592\cdot10^9;\ k_{12}^{(2)} = -0.408\cdot10^9;\ k_{22}^{(2)} = 0.276\cdot10^9$$

and

$$K_{11}^{(1)} = -4.026\cdot10^9;\ K_{12}^{(1)} = 1.842\cdot10^9$$
$$K_{11}^{(2)} = 0.276\cdot10^9;\ K_{12}^{(2)} = 0.408\cdot10^9.$$

Substitution of these into (553) gives

$$\left.\begin{aligned} l_{11} &= -\,0.909;\ l_{12} = 0.56 \\ l_{21} &= \quad 0.416;\ l_{22} = 0.828 \end{aligned}\right\}.$$

The transformation (554), therefore, is

$$\left.\begin{aligned} q_1 &= -\,0.909q_1' + 0.56\ q_2' \\ q_2 &= \quad 0.416q_1' + 0.828q_2' \end{aligned}\right\},$$

and its inverse reads

$$\left.\begin{aligned} q_1' &= -\,0.841q_1 + 0.568q_2 \\ q_2' &= \quad 0.422q_1 + 0.922q_2 \end{aligned}\right\}.$$

Taking the constant voltage $E = 10^3$, we see by inspection that the steady-state values are

$$\left.\begin{aligned} &q_{1s} = \frac{E}{S_1} = 10^{-6};\ q_{2s} = 0 \\ &\dot{q}_{1s} = 0;\ \dot{q}_{2s} = 0 \end{aligned}\right\}.$$

Assuming initial rest conditions, we have for the initial values of the transient charges and currents

$$q_{10} = -\,10^{-6};\ q_{20} = 0;\ \dot{q}_{10} = 0;\ \dot{q}_{20} = 0.$$

Using these in the last transformation above we get for the initial charges and currents in the normal coordinates

$$\left.\begin{aligned} q_{10}' &= 0.841\cdot 10^{-6};\ \dot{q}_{10}' = 0 \\ q_{20}' &= -\,0.422\cdot 10^{-6};\ \dot{q}_{20}' = 0 \end{aligned}\right\}.$$

With the normal mesh charges given by (555) and the currents by the derivatives of these, substitution of the above initial values results in the equations

$$\left.\begin{aligned} N_1 + \overline{N}_1 &= 0.841\cdot 10^{-6} \\ N_1 - \overline{N}_1 &= \quad 0 \end{aligned}\right\};\ \left.\begin{aligned} N_2 + \overline{N}_2 &= -\,0.422\cdot 10^{-6} \\ N_2 - \overline{N}_2 &= \quad 0 \end{aligned}\right\},$$

the solution of which is

$$N_1 = \overline{N}_1 = 0.42\cdot 10^{-6};\ N_2 = \overline{N}_2 = -\,0.211\cdot 10^{-6}.$$

Hence the normal charges and currents are

$$\left.\begin{aligned} q_1' &= 0.841\cdot 10^{-6}\cos g_1 t \\ \dot{q}_1' &= -\,0.841\cdot 10^{-6}\cdot g_1\cdot \sin g_1 t \end{aligned}\right\};\ \left.\begin{aligned} q_2' &= -\,0.422\cdot 10^{-6}\cos g_2 t \\ \dot{q}_2' &= 0.422\cdot 10^{-6}\cdot g_2\cdot \sin g_2 t \end{aligned}\right\}$$

Transforming back to the actual meshes we have after adding the steady-state portions

$$q_1 = -0.764\cdot 10^{-6}\cos g_1 t - 0.236\cdot 10^{-6}\cos g_2 t - 10^{-6}$$
$$q_2 = 0.35\cdot 10^{-6}\cos g_1 t - 0.35\cdot 10^{-6}\cos g_2 t$$

and

$$\dot{q}_1 = \quad 1.035\sin g_1 t + 0.1505\sin g_2 t$$
$$\dot{q}_2 = -0.474\sin g_1 t + 0.233\sin g_2 t$$

When the network is dissipative, the elements of the modular determinant on the mesh-charge basis become instead of (550)

$$k_{ik} = L_{ik}p^2 + R_{ik}p + S_{ik}. \tag{562}$$

If the resistances are small enough so that the network is fully oscillatory, then the roots of the determinantal equation (551) are in the form of conjugate complex pairs. Denoting these by p_ν and $\bar{p}_\nu$ where $\nu = 1, 2, \cdots n$, the transient mesh charges are given by

$$q_k = \sum_{\nu=1,2,\cdots}^{n}(l_{k\nu}N_\nu\cdot e^{p_\nu t} + \bar{l}_{k\nu}\bar{N}_\nu e^{\bar{p}_\nu t}). \tag{563}$$

The direction cosines $l_{k\nu}$ are no longer real. Hence the actual mesh charges can no longer be considered as *geometrical* projections of the normal oscillations q_ν' upon the reference coordinates. The process of projection in this case affects the time phase as well as the amplitude of each oscillation. Although the pictorial significance of the network behavior in terms of normal coordinates may still be retained, the transformation (554) is in general no longer possible.

It is, however, still possible whenever the L, R, and S matrices are *linearly dependent*, namely, if

$$\mathsf{R} = a\mathsf{L} + b\mathsf{S}, \tag{564}$$

where a and b are any real constants. A common case in practice is that for which the resistance and inductance matrices are proportional. We may write this as

$$\mathsf{R} = 2\alpha\mathsf{L}, \tag{565}$$

or more explicitly

$$R_{ik} = 2\alpha L_{ik}, \tag{565a}$$

where α is a positive real constant. Physically this means that all coils in the network have the same resistance-to-inductance ratio. Purely mutual inductance is thus ruled out unless the mutual

branch in question also contains a proportional amount of resistance. Nevertheless a large number of practical cases fall into this category.

When (565a) holds, the k_{ik}'s according to (562) become

$$k_{ik} = L_{ik}p'^2 + S_{ik} \tag{566}$$

where

$$p'^2 = (p^2 + 2\alpha p). \tag{567}$$

The determinantal equation (551) contains only even powers of p'. Considering p'^2 as the variable, the roots must all be *negative* and *real*, as in the non-dissipative case. If we denote these by

$$p_\nu'^2 = -g_{0\nu}^2;\ (\nu = 1, 2, \cdots n), \tag{568}$$

then (567) gives

$$p_\nu^2 + 2\alpha p_\nu + g_{0\nu}^2 = 0, \tag{569}$$

which is identical in form to the determinantal equation for the series R, L, C-circuit. The modes of the network are the roots of (569) for the values (568). We shall consider only the completely oscillatory case defined by

$$\alpha < g_{0\nu}. \tag{570}$$

Then (569) gives

with

$$\left.\begin{aligned} p_\nu &= -\alpha \pm jg_\nu \\ g_\nu &= \sqrt{g_{0\nu}^2 - \alpha^2} \end{aligned}\right\};\ (\nu = 1, 2, \cdots n). \tag{571}$$

The significant feature of this result is that *all the decrements in the network are equal*, and α is this common decrement.

In order to determine the elements (553) of the transformation (556) which will lead to the normal oscillations of our network according to (554), we see that each pair of conjugate modes is defined by one value of $g_{0\nu}$ and hence by means of (566) leads to one set of modular vector components. The n modular vector groups are *real* and are defined by

$$k_{ik}^{(\nu)} = -L_{ik}g_{0\nu}^2 + S_{ik};\ (\nu = 1, 2, \cdots n). \tag{572}$$

With these the minors of the modular determinant $K_{ik}^{(\nu)}$ are formed, and the coefficients $l_{k\nu}$ of the desired transformation are found from (553)

With these coefficients the transformation (556) will simultaneously reduce the inductance, resistance, and elastance matrices to the diagonal

form, i.e., that form which possesses elements on the principal diagonal only. That is

$$\mathsf{A}'\,\mathsf{L}\,\mathsf{A} = \begin{Vmatrix} L_{11}' & 0 & \cdot & 0 \\ 0 & L_{22}' & & 0 \\ \cdot & & & \\ 0 & 0 & \cdot & \cdot\, L_{nn}' \end{Vmatrix}, \tag{573}$$

$$\mathsf{A}'\,\mathsf{R}\,\mathsf{A} = \begin{Vmatrix} R_{11}' & 0 & \cdot\,\cdot & 0 \\ 0 & R_{22}' & \cdot & 0 \\ \cdot & & & \cdot \\ 0 & 0 & \cdot & \cdot\, R_{nn}' \end{Vmatrix}, \tag{574}$$

and

$$\mathsf{A}'\,\mathsf{S}\,\mathsf{A} = \begin{Vmatrix} S_{11}' & 0 & & 0 \\ 0 & S_{22}' & & \cdot\, 0 \\ \cdot & \cdot & & \cdot\;\cdot\;\cdot \\ 0 & 0 & \cdot & \cdot\, S_{nn}' \end{Vmatrix}. \tag{575}$$

The corresponding quadratic forms are reduced to sums of squares

$$T' = \tfrac{1}{2}\sum_{\nu=1}^{n} L_{\nu\nu}'\dot{q}_\nu'^2;\quad F' = \tfrac{1}{2}\sum_{\nu=1}^{n} R_{\nu\nu}'\dot{q}_\nu'^2;\quad V' = \tfrac{1}{2}\sum_{\nu=1}^{n} S_{\nu\nu}'q_\nu'^2. \tag{576}$$

In view of (565*a*) we also have

$$\left.\begin{aligned} R_{\nu\nu}' &= 2\alpha L_{\nu\nu}' \\ F' &= 2\alpha T' \end{aligned}\right\}, \tag{577}$$

and just as in the non-dissipative case

$$S_{\nu\nu}' = g_{0\nu}^2 L_{\nu\nu}. \tag{578}$$

The determination of the constants of integration by transforming to the normal coordinates and evaluating the amplitudes and phases in these separately, proceeds just as in the non-dissipative case. The transformation is *real* and hence the numerical work is readily carried out. In fact this case where the resistance-to-inductance ratio is the same in all the coils is numerically as easily treated as a non-dissipative case[1].

[1] This treatment may readily be extended to the practical case where not only the coils have the same proportionate amount of resistance in series, but all the condensers have a common proportionate amount of leakage conductance in parallel. Thus if we have $G = 2\beta C$ for all the condensers in addition to $R = 2\alpha L$ for all the coils, the elements of the determinant (551) become $k_{ik} = L_{ik}(p + 2\alpha)(p + 2\beta) + S_{ik}$, where the mesh charges are defined operationally in terms of the mesh currents by the relation $i_k = (p + 2\beta)q_k$ in which $p = d/dt$, i.e., on a branch not shared by another mesh, q_k is the accumulated condenser charge (a condenser with its parallel conductance is considered as a single element so that the number of meshes remains the same as before the parallel conductances are added). In place of (567) we let $p'^2 =$

The evaluation of integration constants by transformation to normal coordinates seems to offer an advantage as compared, for example, to the direct evaluation method in that it involves simultaneous systems of only n equations with n unknowns as against $2n$ equations with $2n$ unknowns. With a little additional manipulation, however, the simultaneous system of equations arising in the direct evaluation process[2] may for this same case be separated into two systems of n equations each.

9. The principle of duality.[3] The method of linear transformations as applied to the problem of finding equivalent networks becomes considerably more flexible when it is recognized that it may also be carried out with respect to admittances by making use of the principle of duality with regard to electrical networks. This principle proceeds from a recognition that the following pairs of equations are reciprocal in nature

$$\left.\begin{aligned} e &= Ri \\ i &= Ge \end{aligned}\right\}, \tag{579}$$

and

$$\left.\begin{aligned} e &= L\frac{di}{dt} \\ i &= C\frac{de}{dt} \end{aligned}\right\}. \tag{580}$$

Any quantity which in this sense is the reciprocal of another quantity is said to be the **dual** of that quantity. The equations (579) and (580) show that complete duality exists between voltage, current, and the electric circuit elements. The second equation (579) follows from the

---

$(p + 2\alpha)(p + 2\beta)$ and again find negative real roots $p_\nu'^2 = -g_{0\nu}^2$ from which the $k_{ik}^{(\nu)}$ and hence the direction cosines $l_{k\nu}$ are determined from (553) as before. In place of (569) we now have $p_\nu^2 + 2(\alpha + \beta)p_\nu + 4\alpha\beta + g_{0\nu}^2 = 0$, from which the modes of the network are $p_\nu = -(\alpha + \beta) \pm j\sqrt{g_{0\nu}^2 - (\alpha - \beta)^2}$. Again all the decrements are equal, namely to $(R/2L + G/2C)$. The initial *mesh charges* are determined from the initial charges in the various condensers in the same way as for the case without leakage, namely, from the linear system

$$\sum_{k=1}^{n} S_{ik}q_{k0} = e_{i0};\ (i = 1, 2, \ldots n),$$

in which q_{k0} are the initial mesh charges and e_{i0} the net initial condenser voltage on the contour of the ith mesh.

Incidentally, it is interesting to note that if $\alpha = \beta$, the network impedances $Z(j\omega)$ follow from the corresponding non-dissipative impedances by simply replacing $j\omega$ by $j\omega + 2\alpha$.

[2] See eqs. (441), p. 220, Vol. I.

[3] See Russell, "Alternating Currents."

first by replacing voltage by current, and vice versa, and resistance by conductance. In (580) the second equation follows from the first by interchanging voltage and current and replacing inductance by capacitance. Hence we apparently have

Quantity	*Dual Quantity*
voltage	current
resistance	conductance
inductance	capacitance

The relations (579) are a pair of duals, and so are the equations (580).

The principle of duality is merely a heuristic development of these simple ideas so that they may be extended to networks in general. This extension is greatly assisted by means of the Kirchhoff voltage and current equations which are evidently also a pair of dual relationships. These are

$$\left.\begin{aligned}\sum e &= 0\\ \sum i &= 0\end{aligned}\right\} \tag{581}$$

The first states that the sum of voltages around a closed contour must be zero; and the second that the sum of currents converging toward a branch point must be zero. Since voltages and currents have already been shown to be duals, we see from this that a mesh or closed contour is evidently the dual of a branch point or node.

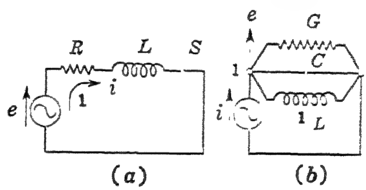


FIG. 89.—The single-mesh circuit and its dual, the single-node circuit.

According to ordinary network ideas, a voltage is considered as the driving force and the resulting current as the response. The dual of this is to consider the current as the driving force and the resulting voltage as the response. This is the interpretation to be given to the second equations (579) and (580).

The dual of a single-mesh circuit containing R, L, and S in series and excited by a voltage source is a single branch-point circuit containing G, C, and $1/L$ in parallel and excited by a current source. These are illustrated in Fig. 89 (*a*) and (*b*) respectively. In the single-mesh case the latter is numbered 1, a positive direction is assigned to the mesh current, and a positive direction is assigned to the driving voltage. In the branch-point case, the branch point is numbered, and positive directions are assigned to the branch-point voltage and the current source. The voltage may here be taken with respect to any arbitrary

datum, but is usually referred to another point in the network, in this case to the return path for the current.

The differential equations for these circuits are respectively

and

$$\left.\begin{aligned} L\frac{di}{dt} + Ri + S\int idt = e \\ C\frac{de}{dt} + Ge + \frac{1}{L}\int edt = i \end{aligned}\right\}. \tag{582}$$

The dual of the general n-mesh network may be formulated by extending these ideas. In this extension it should be kept in mind that the dual of a common branch between two meshes is a common branch between two branch points. Furthermore, whereas the most general common branch between meshes contains resistance, inductance, and elastance in *series*, the most general common branch between branch points contains conductance, capacitance, and reciprocal inductance in *parallel*. Thus a common branch between branch points consists of several parallel paths. This departs somewhat from the definition of a branch as applied to networks on the mesh basis. A similar departure exists in the definition of a branch point. On the mesh basis a branch point is defined as the junction of three or more branches. This is done merely to keep the total number of branches and branch points a minimum. On the branch-point basis, however, it becomes necessary to consider a branch point located at the junction between any two kinds of elements. For example, the single-mesh network of Fig. 89 (a), if viewed in its dual sense, would have to be considered as a three-branch-point network as illustrated in Fig. 90.[1] The voltages of the three branch points may all be considered relative to the return current path. The common branches are in this case not of the most general form since they contain only one kind of element each.

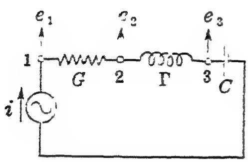


FIG. 90.—Dual representation of the single-mesh circuit in Fig. 89(a).

Just as on the mesh basis it is convenient to introduce elastance in place of capacitance because the series addition of this parameter around a contour thereby becomes simple, so it becomes convenient in the dual representation to introduce a symbol for the reciprocal of inductance in

[1] It should be recognized, of course, that the definition of a *branch* in any network is quite an arbitrary matter and depends entirely upon the requirements of a specific case. Thus a four-terminal network coupling a pair of meshes may be considered symbolically as a single branch if the node potentials and branch currents within it are of no consequence for the problem under consideration.

order that the total parallel inductance at any branch point may be written as the sum of the various reciprocal inductances meeting in that point. We suggest for this purpose the symbol

$$\Gamma = \frac{1}{L}. \tag{583}$$

We may now formulate the dual to the general *n-mesh* network as follows. It is a network consisting of *n branch points* (or *nodes*) and one common datum or return current path with respect to which the various applied current sources are impressed. The branch points and branch-point voltages are correspondingly numbered and reference arrows assigned pointing from the datum to the various branch points. Then if we define the **branch point parameters** C_{rr}, G_{rr}, and Γ_{rr} as the *total* capacitance, conductance, and reciprocal inductance, respectively, meeting in the branch point r; the **mutual parameters** C_{rs}, G_{rs}, and Γ_{rs} as the *negatives* of the capacitance, conductance, and reciprocal inductance, respectively, joining branch points r and s; and the differential-integral operators

$$d_{rs} = C_{rs}\frac{d}{dt} + G_{rs} + \Gamma_{rs}\int dt, \tag{584}$$

then the differential equations expressing the dynamic equilibrium of the dual network become

$$\left.\begin{array}{l} d_{11}e_1 + d_{12}e_2 + \cdots + d_{1n}e_n = i_1 \\ d_{21}e_1 + d_{22}e_2 + \cdots + d_{2n}e_n = i_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ d_{n1}e_1 + d_{n2}e_2 + \cdots + d_{nn}e_n = i_n \end{array}\right\}, \tag{585}$$

in which $i_1, i_2, \ldots i_n$ are the impressed currents at the respective branch points, their positive directions being assumed in the direction from the common datum[1] to the branch points. It should not be thought that this dual network is the equivalent of the general n-mesh network or even its potential equivalent, any more than the network (*b*) of Fig. 89 may be thought of as the equivalent of (*a*). The *dual representation* of the n-mesh network is not the same as the *dual* of that network. For example, (*b*) of Fig. 89 is the dual of (*a*), while Fig. 90 is the dual representation of Fig. 89 (*a*). In passing from the mesh representation to the branch-point representation, the number of branch points in

[1] When all the current sources do not have a *common* return in the datum node, the treatment is given by a superposition of individual solutions for each of which such a common return does exist.

the latter evidently does not equal the number of meshes in the former. The student is cautioned not to become confused on this point.[1]

10. Equivalent networks by means of their dual representation. It should be clear from the above that the behavior of a given network may be expressed either as the current response to impressed voltages or as a voltage response to impressed currents. A two-terminal network is one containing only one source, the terminals of which are the network terminals. For the usual representation the ratio of the driving voltage (assuming now a simple harmonic source with constant amplitude, frequency, and phase) to the resulting current at the terminals is denoted as the driving-point impedance. In the dual representation the ratio of the driving current to the resulting voltage at the terminals is denoted as the driving-point admittance. For a given network this is the reciprocal of the driving-point impedance.

Hence if the dual representation of the network behavior is expressed by

$$\left.\begin{aligned} &\sum_{s=1}^{n} d_{rs}e_s = i_r\ ,\ (r = 1, 2, \cdots n) \\ \text{with}\quad &i_1 \neq 0\ ;\ i_2 = i_3 = \cdots = i_n = 0 \end{aligned}\right\}, \tag{586}$$

where branch point 1 is the driving point, then the linear transformation

$$\left.\begin{aligned} &e_s = \sum_{k=1}^{n} \alpha_{sk}e_k'\ ;\ (s = 1, 2, \cdots n) \\ \text{with}\quad &\alpha_{11} = 1;\ \alpha_{12} = \alpha_{13} = \cdots = \alpha_{1n} = 0 \end{aligned}\right\} \tag{587}$$

[1] It is significant to note in connection with the dual method of expressing the network equilibrium that this approach to the analysis of network behavior does not necessarily require that the sources be defined as *current* functions, as might be inferred from a casual review of this matter. For example, let us assume in (585) that only the one source i_1 is present and that $i_2 = i_3 = \cdots = i_n = 0$, i.e., no sources are applied to the remaining nodes. If this one source is specified in terms of voltage instead of current, then the node potential e_1 is that known voltage. We may then disregard the first equation (585), the remaining $n - 1$ equations, after the terms with e_1 are transposed to the right-hand sides, may be solved uniquely for the remaining $n - 1$ node potentials $e_2 \ldots e_n$, and the desired network behavior is thus obtained. If μ equals the *total* number of nodes and σ the number of branches in the network, this leads to $\mu - 2$ simultaneous equations (since one of the node potentials is that of the known source and another is taken as the datum). Recalling the relation (272), p. 129, of Vol. I, the number of independent mesh currents (if the equilibrium were expressed in terms of these) are for this network $\sigma - \mu + 1$, and the solution on this basis, therefore, involves this number of simultaneous equations and unknowns. The choice between the *mesh* or the *node* bases in a given instance depends upon which involves the least number of unknowns. Recognition of this may greatly simplify the treatment in some cases.

will lead to new networks having the same driving-point admittance function. Denoting the capacitance, conductance, and reciprocal inductance matrices of the given network by[1]

$$\left.\begin{array}{l} \mathsf{C} = \| C_{rs} \| \\ \mathsf{G} = \| G_{rs} \| \\ \mathsf{\Gamma} = \| \Gamma_{rs} \| \end{array}\right\} \tag{588}$$

respectively, and the transformation matrix by

$$\mathsf{A} = \left\| \begin{array}{ccc} \alpha_{11} & \cdots & \alpha_{1n} \\ \cdot & \cdots & \cdot \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{array} \right\|, \tag{589}$$

the matrices of the equivalent network are

$$\left.\begin{array}{l} \mathsf{C}' = \mathsf{A}'\,\mathsf{C}\,\mathsf{A} \\ \mathsf{G}' = \mathsf{A}'\,\mathsf{G}\,\mathsf{A} \\ \mathsf{\Gamma}' = \mathsf{A}'\,\mathsf{\Gamma}\,\mathsf{A} \end{array}\right\}, \tag{590}$$

where A' is the conjugate of A as previously pointed out.

Equivalent networks obtained by this method form a group in which the number of branch points is invariant. The transformations on the mesh basis form a group in which the number of meshes is invariant.[2] These groups are in general distinct; i.e., it is not generally possible to obtain the networks of one group from those of the other by transformations on the basis of the other group. Another way of putting this is to say that the equivalent networks obtained on the branch-point basis cannot in general be also obtained on the mesh basis.

As an illustration consider the single series R, L, S-network of Fig. 89(a) whose dual representation is shown in Fig. 90. The latter is characterized by the matrices

$$\mathsf{C} = \left\| \begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C \end{array} \right\|; \quad \mathsf{G} = \left\| \begin{array}{ccc} G & -G & 0 \\ -G & G & 0 \\ 0 & 0 & 0 \end{array} \right\|; \quad \mathsf{\Gamma} = \left\| \begin{array}{ccc} 0 & 0 & 0 \\ 0 & \Gamma & -\Gamma \\ 0 & -\Gamma & \Gamma \end{array} \right\|.$$

[1] These are the matrices of the quadratic forms representing the stored and loss energy functions for the network on the dual basis. These are:

$$V = \tfrac{1}{2}\Sigma\, C_{rs} e_r e_s$$
$$T = \tfrac{1}{2}\Sigma \Gamma_{rs} \varphi_r \varphi_s$$
$$F = \tfrac{1}{2}\Sigma\, G_{rs} e_r e_s$$

where the φ_r's are defined by

$$\varphi_r = \int e_r dt; \text{ or } e_r = \frac{d\varphi_r}{dt}.$$

[2] These statements are subject to certain restrictions to be pointed out later.

From what has been said in section 5 above, we can, for example, obtain an equivalent driving-point admittance by multiplying any but the first rows and columns by a factor. Suppose we multiply the second rows and columns by 2. Then we have

$$\mathsf{C}' = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C \end{Vmatrix}; \quad \mathsf{G}' = \begin{Vmatrix} G & -2G & 0 \\ -2G & 4G & 0 \\ 0 & 0 & 0 \end{Vmatrix}; \quad \Gamma' = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & 4\Gamma & -2\Gamma \\ 0 & -2\Gamma & \Gamma \end{Vmatrix}$$

The corresponding equivalent driving-point admittance network is shown in Fig. 91. It contains a negative resistance and a negative inductance which may be got rid of by the use of ideal transformers. The resulting network of Fig. 91 may be impractical, but it illustrates the possibilities of the method nevertheless. It is clear that, although the number of branch points is the same, the number of meshes certainly is not. Hence it is clear that this equivalent network could not have been got on the mesh basis.

The C, G, and Γ matrices which are obtained by the branch-point representation also have positive definite quadratic forms when got from a physical network, and this property is not destroyed by any real non-singular transformation.

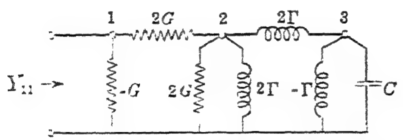


FIG. 91.—Network whose driving-point impedance is the same as that for the single-mesh circuit of Fig. 89(*a*) or Fig. 90.

It might also be pointed out here that it is possible on either the mesh or branch-point basis to obtain equivalent networks with more meshes or branch points than the original. If the given set of three matrices in either case are augmented by inserting at corresponding points on their principal diagonals arbitrary elements and then filling out the remaining positions of these rows and columns by zeros, the impedance function will be unaffected because this process corresponds to adding a mesh or branch point having no coupling with any other mesh or branch point. The augmented matrices may then be transformed to find equivalent networks in the usual fashion.

11. Inverse networks. In section 8 of the preceding chapter we discussed a method of finding an inverse impedance when the given impedance is in the form of a ladder structure with components of the same form. We are now in a position to solve this problem generally.[1]

Suppose we have an *n-mesh* network with the driving-point impedance Z_1 and the inductance, resistance, and elastance matrices L, R,

[1] This was suggested to the writer by O. Brune.

and S. If we construct an *n-branch-point* network with the driving-point admittance Y_2 and the capacitance, conductance, and reciprocal inductance matrices C, G, and Γ, then we will have

$$Z_1 = Y_2 \tag{591}$$

provided

$$\mathsf{C} = \mathsf{L};\ \mathsf{G} = \mathsf{R};\ \Gamma = \mathsf{S}. \tag{592}$$

If we denote the driving-point impedance of the second network by

$$Z_2 = Y_2^{-1},$$

then (591) reads

$$Z_1 Z_2 = 1, \tag{591a}$$

and hence (592) is the necessary and sufficient condition that the networks be each other's inverses.

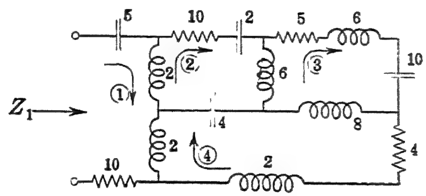


FIG. 92.—Two-terminal network represented on the mesh basis.

If we want

$$Z_1 Z_2 = R^2, \tag{591b}$$

then the conditions (592) become

$$R^2\mathsf{C} = \mathsf{L};\ R^2\mathsf{G} = \mathsf{R};\ R^2\Gamma = \mathsf{S}. \tag{592a}$$

The quantity R is evidently restricted to real values. These conditions read more explicitly

$$R^2C_{rs} = L_{rs};\ R^2G_{rs} = R_{rs};\ R^2\Gamma_{rs} = S_{rs}. \tag{592b}$$

To illustrate this we consider the four-mesh network of Fig. 92. The indicated values are in *henries*, *ohms*, and *reciprocal farads*. Writing the inductance, resistance, and elastance matrices by inspection and applying (592a) we have

$$\mathsf{L} = \left\| \begin{matrix} 4 & -2 & 0 & -2 \\ -2 & 8 & -6 & 0 \\ 0 & -6 & 20 & -8 \\ -2 & 0 & -8 & 12 \end{matrix} \right\| = R^2\mathsf{C},$$

$$\mathsf{R} = \begin{Vmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{Vmatrix} = R^2\mathsf{G},$$

and

$$\mathsf{S} = \begin{Vmatrix} 5 & 0 & 0 & 0 \\ 0 & 6 & 0 & -4 \\ 0 & 0 & 10 & 0 \\ 0 & -4 & 0 & 4 \end{Vmatrix} = R^2\Gamma.$$

Letting $R = 1$ for numerical simplicity, the inverse network is constructed directly from these matrices and is shown in Fig. 93. The

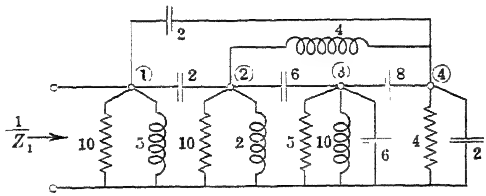


FIG. 93.—Dual network of the one shown in Fig. 92. The driving-point impedances of these networks are inverse.

values are in *farads*, *reciprocal ohms*, and *reciprocal henries*. The branch points are numbered ①, ②, ③, ④.

PROBLEMS TO CHAPTER VI

6-1. A driving-point reactance network consists of two resonant components in parallel with:

$$L_1 = 2 \cdot 10^{-3};\ C_1 = 2 \cdot 10^{-6};\ L_3 = 0.5 \cdot 10^{-3};\ C_3 = 4 \cdot 10^{-6}.$$

Using the node basis of network representation, set down the corresponding capacitance and reciprocal inductance matrices, and find the proper transformation matrices which will yield the three equivalent canonic forms. Determine the parameter values in the latter, and check by carrying through the same transformations by the Foster and Cauer methods.

6-2. Consider the driving-point reactance consisting of three resonant components in parallel with

$$L_1 = 9 \cdot 10^{-3};\ C_1 = 9 \cdot 10^{-6},\ L_2 = 4 \cdot 10^{-3},\ C_2 = 4 \cdot 10^{-6};\ L_3 = 10^{-3};\ C_3 = 10^{-6},$$

and find the three canonic equivalent forms by linear transformations on the mesh basis.

6-3. Repeat Problem 6-2 on the node basis and check the results with those of Problem 6-2 as well as with an independent determination using the Foster and Cauer methods.

6–4. Consider the inductance and elastance matrices for either of the Cauer networks of Problem 6–1, and find the linear transformations which will simultaneously transform the inductance and elastance matrices to the diagonal (normal) form Show that the normal form for the network corresponds to that consisting of the two resonant components in parallel, thus physically realizing the normal coordinates as actual meshes

6–5. Repeat Problem 6–4 for the network defined in Problem 6–2.

6–6. In the Problem 6–4 assume that the coils and condensers are dissipative and have the R/L and G/C ratios 100 and 1, respectively. Find the common damping constant for the network. Determine the network response in both meshes for a suddenly applied constant emf of 1000 volts and initial rest conditions in the case of either of the Cauer forms by transformation to the normal form and then transforming back again. Show that the response at the input terminals is the same for the four equivalent canonic forms.

6–7. On the basis of the results of Problem 6–6 show that the Foster and Cauer reactance theorems may be extended without change in their manner of application to any uniformly dissipative network, i.e., one in which all R/L ratios are equal and all G/C ratios are equal.

6–8. Starting with an n-mesh network, show that equivalents with more meshes may in general be obtained by representing the given network on the node basis and subjecting this to a linear transformation which keeps the number of independent nodes invariant.

6–9. A network which is driven from mesh 1 alone has the following resistance, inductance, and elastance matrices:

$$\mathsf{R} = \begin{Vmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{Vmatrix}; \quad \mathsf{L} = 10^{-3} \cdot \begin{Vmatrix} 11 & -1 & 0 & -10 \\ -1 & 5 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ -10 & 0 & 0 & 17 \end{Vmatrix};$$

$$\mathsf{S} = 10^{6} \cdot \begin{Vmatrix} 2 & 0 & -2 & 0 \\ 0 & 5 & -5 & 0 \\ -2 & -5 & 7 & 0 \\ 0 & 0 & 0 & 9 \end{Vmatrix}.$$

Sketch the network. Determine the structure of, as well as the element values in, the network whose driving-point impedance is the inverse of that of the original network with respect to the constant value 500.

CHAPTER VII

ARTIFICIAL AND LUMP-LOADED LINES

1. Lumped-section equivalents to the long line. It is frequently desirable for investigatory purposes to be able to construct a lumped four-terminal network whose external behavior will simulate that of a given line or cable. Such a network is referred to as an **artificial line.** In the first part of this chapter we wish to discuss means whereby such networks may be physically realized to within a sufficient degree of approximation for any given case.

In section 10 of Chapter IV we discussed three common types of lumped structures, namely, the symmetrical T-, Π-, and lattice-structures, and determined the relations between the component impedances of these structures and their corresponding propagation and characteristic impedance functions. These relations may be used either for the determination of these functions when the structures themselves are given, or to determine the component impedances of the structures when propagation and characteristic impedance functions are prescribed. This last alternative immediately suggests a form of solution to our present problem, namely, to use one of these structures with component impedances determined in agreement with the propagation and characteristic impedance functions of the given line or cable. Let us examine the possibilities of this process more carefully.

According to the notation adopted in Chapter II, the propagation and characteristic impedance functions of a line l miles in length are given by

$$\left.\begin{aligned} \alpha l &= l\sqrt{(R + jL\omega)(G + jC\omega)} \\ \text{and} \qquad Z_0 &= \sqrt{\frac{R + jL\omega}{G + jC\omega}} \end{aligned}\right\}, \tag{593}$$

where R, L, G, C are the line parameters per mile and ω is the angular frequency at which the line is excited in the steady state. Using equations (426) and (429) of Chapter IV for the symmetrical T- and Π-structures of Fig. 46, we find that the component impedances and admittances, respectively, must be

$$\left.\begin{aligned} z_1 &= 2Z_0 \tanh \frac{\alpha l}{2} \\ z_2 &= \frac{Z_0}{\sinh \alpha l} \end{aligned}\right\}, \tag{594}$$

and

$$\left.\begin{aligned} y_1 &= \frac{1}{Z_0 \sinh \alpha l} \\ y_2 &= \frac{2 \tanh \dfrac{\alpha l}{2}}{Z_0} \end{aligned}\right\}. \tag{595}$$

At a single frequency these component impedances or admittances are easily realizable by means of the ordinary circuit elements. This needs no further discussion. In communication work it is necessary to find structures which will simulate the given line over a prescribed *band* of frequencies. This is quite another matter. Recalling the discussion of α and Z_0 in Chapter III, we recognize that the relations (594) and (595) represent anything but simple functions of frequency.

There is, however, one fairly obvious clue to this situation which we notice upon further study of the above analytic results. Namely, suppose for the moment that the length of line which we are desirous of simulating is extremely short. Over a fixed frequency range this may make the arguments of the hyperbolic functions small enough so that the latter might be replaced by the arguments themselves without excessive error. Analytically we have for $l \rightarrow 0$

$$\left.\begin{aligned} z_1 &\rightarrow Z_0\alpha l = (R + jL\omega)l \\ z_2 &\rightarrow \frac{Z_0}{\alpha l} = \frac{1}{(G + jC\omega)l} \end{aligned}\right\}, \tag{594a}$$

and

$$\left.\begin{aligned} y_1 &\rightarrow \frac{1}{Z_0\alpha l} = \frac{1}{(R + jL\omega)l} \\ y_2 &\rightarrow \frac{\alpha l}{Z_0} = (G + jC\omega)l \end{aligned}\right\}. \tag{595a}$$

These are certainly of a physically realizable form. The resulting T- or Π-structures will be the desired lumped equivalents for the line over a region of frequency for which these approximations are justified. Since α becomes a linear function of frequency for large frequencies, we expect that the above procedure will be justified only up to a certain frequency beyond which the error exceeds a specified maximum. Regarding the short line length to which the process is restricted, this can evidently be extended by cascading a sufficient number of structures

of this kind. Examination of (594a) or (595a) shows that such a cascade of structures, if carried to the limit $l \to 0$ (for which the number of structures becomes infinite), leads to the uniform line itself. Thus this procedure could have been justified on a heuristic basis from the beginning.

The practical solution of our problem on this basis requires that we be able to determine the number of sections and the parameter values of each for a given finite length of line, frequency range, and maximum allowable error. Since there is no appreciable difference between the use of T- or Π-structures, we shall restrict the following discussion to the T. The student may carry through the same argument for the Π-structure as an exercise.

We begin by turning our problem about, and suppose that we have before us a cascade of n symmetrical T-structures, for each of which

$$\left.\begin{aligned} z_1 &= \frac{l}{n}(R + jL\omega) \\ \frac{1}{z_2} &= \frac{l}{n}(G + jC\omega) \end{aligned}\right\}, \tag{596}$$

and ask ourselves: To within what maximum error and over what frequency range does this cascade approximate the uniform line for which α and Z_0 are given by (593)? Using equations (424) and (425a) of Chapter IV, and also the fact brought out in section 8 of that chapter that the propagation function of n structures in cascade equals n times that per structure and that the characteristic impedance is independent of the number of cascaded structures, we have for the characteristic impedance and propagation functions of our n T-sections in cascade

$$Z_T = \sqrt{z_1 z_2}\sqrt{1 + \frac{z_1}{4z_2}}; \tag{597}$$

and

$$n\gamma = 2n \sinh^{-1}\sqrt{\frac{z_1}{4z_2}}. \tag{598}$$

Using (593) and (596) these become

$$Z_T = Z_0\sqrt{1 + \left(\frac{\alpha l}{2n}\right)^2}, \tag{597a}$$

and

$$n\gamma = 2n \sinh^{-1}\left(\frac{\alpha l}{2n}\right). \tag{598a}$$

These expressions enable us to determine the errors for any given

case. This becomes more evident if we use expansions for the functions (597a) and (598a), assuming that

$$\left(\frac{\alpha l}{2n}\right) \ll 1. \tag{599}$$

For the radical we can use the binomial expansion. For the inverse hyperbolic sine we have

$$\sinh^{-1}x = x - \frac{1}{2}\cdot\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\cdot\frac{x^5}{5} - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\cdot\frac{x^7}{7} + \cdots.$$

Thus we have

$$Z_T = Z_0\left\{1 + \frac{1}{2}\left(\frac{\alpha l}{2n}\right)^2 - \frac{1}{8}\left(\frac{\alpha l}{2n}\right)^4 + \cdots\right\}, \tag{600}$$

and

$$n\gamma = \alpha l\left\{1 - \frac{1}{6}\left(\frac{\alpha l}{2n}\right)^2 + \frac{3}{40}\left(\frac{\alpha l}{2n}\right)^4 - \cdots\right\}. \tag{601}$$

If we write

$$\left.\begin{aligned} Z_T &= Z_0(1 + \delta_z) \\ n\gamma &= \alpha l(1 + \delta_\alpha) \end{aligned}\right\}, \tag{602}$$

where δ_z and δ_α are the decimal errors in Z_T and $n\gamma$ respectively, then these are approximately given by

$$\left.\begin{aligned} \delta_z = -3\delta_\alpha &= \frac{1}{8}\cdot\left(\frac{l}{n}\right)^2\cdot\alpha^2 \\ &= \frac{1}{8}\cdot\left(\frac{l}{n}\right)^2\cdot(R + jL\omega)(G + jC\omega) \end{aligned}\right\}. \tag{603}$$

The error in Z_T is of opposite sign and three times as large as in $n\gamma$. For a given ratio of (l/n) it increases with frequency, being smallest at zero frequency, while for a given frequency it increases as the square of the line length and decreases as the reciprocal of the square of the number of cascaded sections. The highest essential frequency for which the artificial line is to be used will, therefore, determine the number of sections needed for a given case when the maximum allowable errors are specified. When the maximum errors in Z_T and $n\gamma$ are specified of the same magnitude, then δ_z will evidently govern the design.

Since at high frequencies, where the error is largest, the dissipation has little effect, we can write (603) in an approximate form sufficient for most practical cases by neglecting R and G. Thus

$$\delta_z \cong -\frac{LC\omega^2}{8}\left(\frac{l}{n}\right)^2. \tag{603a}$$

Since by equation (27) of Chapter I

$$LC \cong \varepsilon\mu$$

where ε and μ are the dielectric constant and permeability of the medium surrounding the line conductors, we see that the error will be larger for paper-insulated cables than for aerial lines, other factors being equal.

Denoting the highest essential angular frequency by ω_c, the expression (603a) may be used to solve for the number of sections when the maximum allowable error is specified. For this we have if we merely consider the magnitude of the error and not its sign

$$n = \frac{l\omega_c\sqrt{LC}}{2\sqrt{2\delta_z}}. \tag{604}$$

The component impedances for each T-section are then given by (596).

This may be expressed in another way if in (603) we again neglect dissipation and use

$$\alpha_2 = \frac{2\pi}{\lambda}.$$

Defining

$$n_\lambda = \frac{n\lambda}{l} = \text{number of sections per wave length,} \tag{605}$$

then

$$\delta_z = -\frac{\pi^2}{2n_\lambda^2}, \tag{603b}$$

and

$$n_\lambda = \frac{\pi}{\sqrt{2\delta_z}}, \tag{604a}$$

without regard for the sign of δ_z. The shortest essential wave length will govern the design.

2. The use of lattice structures. There are numerous other ways of constructing artificial lines. The lattice structure which we have already mentioned in this connection offers a number of possibilities. According to equation (432) of Chapter IV, the component impedances of a symmetrical lattice having αl and Z_0 as its propagation and characteristic impedance functions are

$$\left.\begin{aligned} Z_a &= Z_0 \tanh\frac{\alpha l}{2} \\ Z_b &= Z_0 \coth\frac{\alpha l}{2} \end{aligned}\right\}. \tag{606}$$

The procedure from this point on is quite different from that adopted

for T- and Π-section artificial lines. There we had no means of physically realizing both component impedances directly. Here, however, such a procedure is possible; in fact, it is possible in a number of ways.

Since

$$Z_a Z_b = Z_0^2, \tag{607}$$

we see that Z_b is the reciprocal of Z_a with respect to Z_0^2. Hence as soon as Z_a is realized we can find Z_b by reciprocation.[1] We can, therefore, confine our attention to the realization of Z_a. On account of the fact that the characteristic impedance of the symmetrical lattice is equal to the square root of $Z_a Z_b$, we see that this artificial line will have the correct Z_0 no matter how approximate Z_a and Z_b may be. All the error will be in the propagation function.

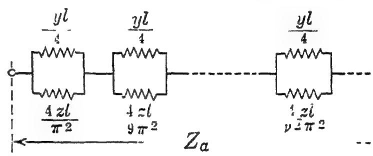


FIG. 94.—Lattice impedance according to the expansion (610).

For the physical realization of Z_a according to the first equation (606), we may proceed in a manner which is very similar to Foster's method of realizing a reactance function. Namely, we may expand either Z_a or $1/Z_a$ into partial fractions by making use of the partial fraction expansions of the *tanh* and *coth* functions. These are

$$\tanh x = \sum_{\nu=1,3,5\ldots}^{\infty} \frac{8x}{\nu^2\pi^2 + 4x^2}, \tag{608}$$

and

$$\coth x = \frac{1}{x} + \sum_{\nu=1,2,3\ldots}^{\infty} \frac{2x}{\nu^2\pi^2 + x^2}. \tag{609}$$

Using these together with (593) and

$$y = (G + jC\omega);\ z = (R + jL\omega),$$

we have

$$Z_a = \sum_{\nu=1,3,5\ldots}^{\infty} \frac{4zl}{\nu^2\pi^2 + yzl^2}, \tag{610}$$

and

$$\frac{1}{Z_a} = \frac{2}{zl} + \sum_{\nu=1,2,3\ldots}^{\infty} \frac{4yl}{4\nu^2\pi^2 + yzl^2}. \tag{611}$$

[1] Although the process of reciprocation as discussed in section 8 of Chapter V is in general possible only with respect to a real constant (resistance), there are instances, such as the present, in which the method is not so restricted. This the reader will readily recognize from the following detailed discussion.

The first of these we recognize as the impedance of the structure shown in Fig. 94; the second is the admittance of the structure shown in Fig. 95. These are similar to Foster's anti-resonant components in series and resonant components in parallel respectively. The student may reciprocate these structures with respect to Z_0^2 to find Z_b.

Another form in which Z_a can be realized is similar to Cauer's extension of the Foster theorem. This is obtained by expanding the tanh function in a continued fraction, thus

$$\tanh x = \frac{1 \,|}{|1/x} + \frac{1 \,|}{|3/x} + \frac{1 \,|}{|5/x} + \frac{1 \,|}{|7/x} + \cdots. \tag{612}$$

For Z_a this gives

$$Z_a = \frac{1 \,|}{|2/zl} + \frac{1 \,|}{|6/yl} + \frac{1 \,|}{|10/zl} + \frac{1 \,|}{|14/yl} + \cdots. \tag{613}$$

This is the impedance of the infinite ladder network shown in Fig. 96.

These three forms for the exact realization of Z_a are all theoretically infinite networks. Practically, approximations will have to be made by using only a finite portion, the error becoming smaller with each added branch. Just how the error in Z_a is linked with the number of branches to which these networks are built out is rather difficult to express analytically. Since the error is also a function of frequency, the maximum error would have to be determined and expressed in terms of the extent of the network. The error in the propagation function would then have to be correlated with the error in Z_a by means of equation (431) of Chapter IV, which for our case reads

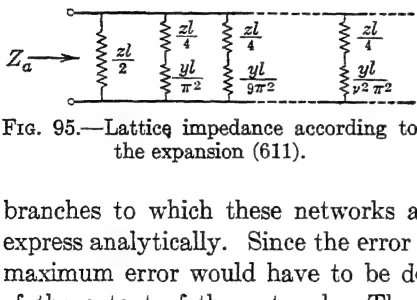


Fig. 95.—Lattice impedance according to the expansion (611).

$$\gamma = 2 \tanh^{-1}\left(\frac{Z_a}{Z_b}\right). \tag{614}$$

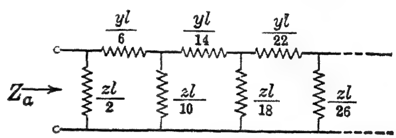


Fig. 96.—Lattice impedance according to the expansion (613).

This correlation may, however, be simply given if we approach our problem in the following indirect manner.

It is possible to design an artificial line on the basis of a lattice structure in a more satisfactory and perhaps more economical manner if we begin with the approximation of neglecting dissipation, with the idea of bringing it into consideration subsequently. If we do this, then

the component impedances (606) become the following reactance functions

$$\left.\begin{aligned} Z_a &= j\sqrt{\frac{L}{C}}\tan\sqrt{LC}\cdot\frac{\omega l}{2} \\ Z_b &= -j\sqrt{\frac{L}{C}}\cot\sqrt{LC}\cdot\frac{\omega l}{2} \end{aligned}\right\}. \tag{615}$$

Here Z_a, for example, has alternate zeros and poles (beginning with a zero) uniformly spaced from the origin to infinity. Up to a given top frequency, only a finite number of critical frequencies are involved, so that it is logical to attempt an approximation to Z_a by means of the Foster or Cauer reactance theorems.

This approximation to Z_a, which we shall denote by Z_a', may evidently be written in the form

$$Z_a' = H\frac{jx\left(1-\frac{x^2}{x_2^2}\right)\left(1-\frac{x^2}{x_4^2}\right)\cdots\left(1-\frac{x^2}{x_n^2}\right)}{\left(1-\frac{x^2}{x_1^2}\right)\left(1-\frac{x^2}{x_3^2}\right)\cdots\left(1-\frac{x^2}{x_{n-1}^2}\right)}, \tag{616}$$

in which the variable x is proportional to frequency, and the parameters $x_1, x_2, \ldots, x_n$ represent the uniformly spaced critical frequencies. In (616) the last of these is a zero of Z_a' since the expression is written for *n-even*. For *n-odd* the form of Z_a' is the same except that the last factor represents a pole and hence is in the denominator instead of in the numerator. In any case n equals the total number of non-zero critical frequencies, and H is a constant multiplier.

If we let

$$x = \frac{\omega l\sqrt{LC}}{\pi}, \tag{617}$$

then Z_a may be put into a similar form by using the infinite product expansion of the tangent function. We thus have

$$\begin{aligned} Z_a &= \frac{j\pi}{2}\sqrt{\frac{L}{C}}\prod_{s=1}^{\infty}\frac{x\left(1-\frac{x^2}{4s^2}\right)}{\left(1-\frac{x^2}{(2s-1)^2}\right)} \\ &= \frac{j\pi}{2}\sqrt{\frac{L}{C}}\cdot\frac{x\left(1-\frac{x^2}{2^2}\right)\left(1-\frac{x^2}{4^2}\right)\cdots}{\left(1-x^2\right)\left(1-\frac{x^2}{3^2}\right)\cdots}. \end{aligned} \tag{615a}$$

By comparison of this with (616) and (617) we see that the parameters x_ν must have the form

$$x_\nu = \frac{\omega_\nu l\sqrt{LC}}{\pi} = \nu; \; (\nu = 1, 2, 3 \cdots n), \tag{617a}$$

where ω_ν are the angular critical frequencies

$$\omega_\nu = \frac{\nu\pi}{l\sqrt{LC}}. \tag{617b}$$

In order to establish a criterion for the degree to which Z_a' approximates Z_a, we shall study the ratio Z_a/Z_a' over the range $0 \leq x \leq n + 1$. The highest corresponding frequency follows from (617b) and is

$$\omega_{n+1} = \frac{\pi(n+1)}{l\sqrt{LC}}. \tag{617c}$$

This we shall call the *approximation range*. By means of (616), (615a), and (617a) this ratio is given by the infinite product

$$\frac{Z_a}{Z_a'} = \frac{\pi}{2H}\sqrt{\frac{L}{C}} \cdot \frac{\left(1 - \frac{x^2}{(n+2)^2}\right)\left(1 - \frac{x^2}{(n+4)^2}\right) \cdots}{\left(1 - \frac{x^2}{(n+1)^2}\right)\left(1 - \frac{x^2}{(n+3)^2}\right) \cdots}. \tag{618}$$

Since the first n factors in Z_a cancel with those in Z_a', the ratio is a continuous function of x within the approximation range, i.e., it has no critical frequencies there. Our immediate object is to make a plot of the ratio within this range.

On account of the slow convergence of (618) it is not very suitable for purposes of calculation. Knowing that the ratio is a continuous function of x for $0 \leq x \leq n + 1$, we can overcome this difficulty by using, instead of (618), the ratio formed by the finite expression (616) and the tangent function itself so long as we avoid the zeros and poles as points at which to calculate the numerical value of the ratio.[1] For the sake

[1] At the zeros and poles the ratio becomes indeterminate, but may be evaluated, of course, in the usual manner. Denoting these points by $x = \nu$ we find for ν-even

$$\left(\frac{Z_a}{Z_a'}\right) = \frac{\pi}{4j}\sqrt{\frac{L}{C}}\left\{x\left(1 - \frac{x^2}{x_\nu^2}\right) Y_a'\right\}_{x=\nu},$$

and for ν-odd

$$\left(\frac{Z_a}{Z_a'}\right)^{-1} = \frac{\pi}{4j}\sqrt{\frac{C}{L}}\left\{x\left(1 - \frac{x^2}{x_\nu^2}\right) Z_a'\right\}_{x=\nu},$$

where $Y_a' = 1/Z_a'$.

of simplicity we choose those points at which the magnitude of the tangent equals unity. Since by (615) and (617)

$$Z_a = j\sqrt{\frac{L}{C}} \tan \frac{\pi}{2}x, \tag{615b}$$

these correspond to

$$x = r; \quad \left(r = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots \frac{2n+1}{2}\right) \tag{619}$$

At these points the ratio then becomes

$$\frac{Z_a}{Z_a'} = H^{-1}\sqrt{\frac{L}{C}}\left|\frac{(1-r^2)\left(1-\frac{r^2}{3^2}\right)\left(1-\frac{r^2}{5^2}\right)\cdots\left(1-\frac{r^2}{(n-1)^2}\right)}{r\left(1-\frac{r^2}{2^2}\right)\left(1-\frac{r^2}{4^2}\right)\cdots\left(1-\frac{r^2}{n^2}\right)}\right|. \tag{618a}$$

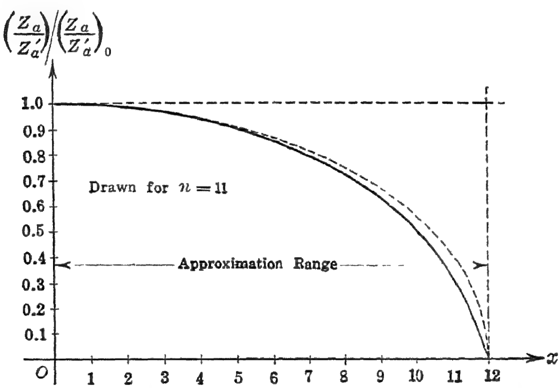


FIG. 97.—Graphical representation of the ratio (620) showing the approximation obtainable in an artificial line design for which the lattice reactance contains eleven coincident critical frequencies only.

The discussion is simplified somewhat if we normalize this ratio by dividing by its value at $x = 0$ which, by (618), is evidently

$$\left(\frac{Z_a}{Z_a'}\right)_0 = \frac{\pi}{2H}\sqrt{\frac{L}{C}}. \tag{618b}$$

The normalized ratio then has the value unity at $x = 0$, while at those points defined by (619) it is given by

$$\left(\frac{Z_a}{Z_a'}\right)\Big/\left(\frac{Z_a}{Z_a'}\right)_0 = \frac{2}{\pi}\left|\frac{(1-r^2)\left(1-\frac{r^2}{3^2}\right)\left(1-\frac{r^2}{5^2}\right)\cdots\left(1-\frac{r^2}{(n-1)^2}\right)}{r\left(1-\frac{r^2}{2^2}\right)\left(1-\frac{r^2}{4^2}\right)\cdots\left(1-\frac{r^2}{n^2}\right)}\right|. \tag{620}$$

This form is quite convenient for purposes of calculation so long as n is not too large. If n is odd the result is the same except that the last factor occurs in the numerator. The solid curve of Fig. 97 shows a plot for the case $n = 11$. The degree to which unity is approximated is evidently not very good except over a small fraction of the range $0 \leq x \leq n + 1$.[1]

This situation may be very materially improved without making the network Z_a' any more extensive by moving the last pole of this reactance function some distance beyond the corresponding pole of the tangent function. The result of such a process is illustrated in Fig. 98.

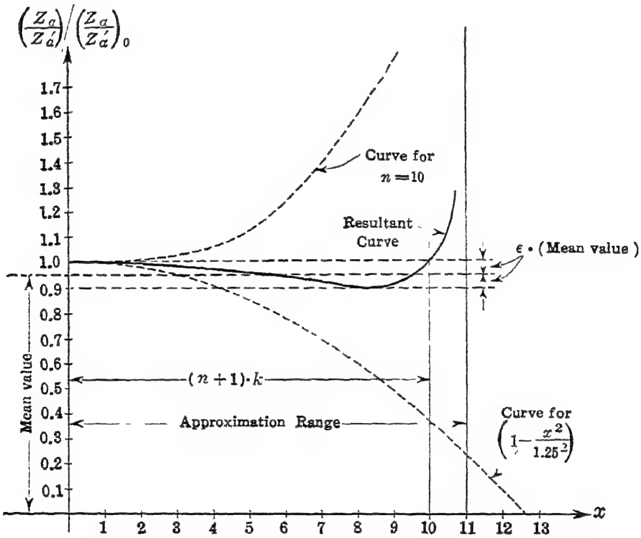


FIG. 98.—Approximation obtained in an artificial line design for which the lattice reactance has ten coincident and one non-coincident critical frequency.

Here the upper dotted curve represents a plot of the normalized ratio for $n = 10$, i.e., for the case where Z_a' contains a total of ten poles and zeros which coincide with the first ten of the tangent function. The lower dotted curve is for one additional pole in Z_a' located at $x = 12.5$. The resultant normalized ratio, which is given by the product of the two dotted curves, may be seen to approximate a constant value very much better than the solid curve of Fig. 97.

It is sometimes convenient to express the degree of approximation in terms of a tolerance and coverage. The tolerance (denoted by ϵ) is the

[1] The reader should note that so far this method of design leads to identical results as those obtained from the expansions (610), (611), or (612) when dissipation is neglected.

ratio of the maximum deviation of the function from its mean value to the mean value as indicated in the figure. The coverage (denoted by k) is the ratio of that region within which the deviation ϵ is not exceeded to the width of the approximation range.[1] In the above example $\epsilon = 0.05$ and $k = 0.9$. By moving the last pole of Z_a' beyond the point $x = 12.5$, the tolerance may be decreased, but the coverage will also be decreased. If the last pole is moved to the right a sufficient distance, the minimum in the resultant curve may be altogether suppressed. Fig. 98 thus illustrates only one of a variety of results obtainable by this process.

It is interesting to note that the idea involved here may be continued with a resultant further improvement in the degree of approximation. For example, if we again let the first ten poles and zeros of Z_a' coincide with those of the tangent function, and then introduce a pole at $x = 11.2$ and a final zero at $x = 15.8$, we obtain the resultant curve shown in Fig. 98a for which the tolerance is less than 2 per cent and the coverage about 94 per cent with a mean value equal to unity. By

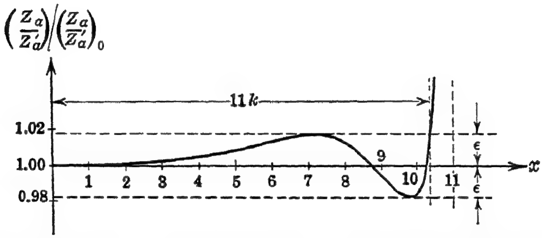


FIG. 98a.—Approximation obtained from ten coincident and two non-coincident critical frequencies in Z_a'.

moving the additional pole and zero farther to the right, the oscillatory character of the curve may in this case also be suppressed.

The locations of the non-coincident poles and zeros in Z_a' may be determined in a specific problem by means of a cut-and-try process entirely. The method involved here, however, has much in common with that discussed in Chapter X (and also to some extent in Chapter

[1] The constant multiplier H in the expression for Z_a' is thus determined from (618b) by means of the relation

$$\frac{2H}{\pi}\sqrt{\frac{C}{L}} = \text{“mean value.”}$$

In Fig. 98 the "mean value" is chosen so that the origin becomes a point of maximum deviation. By setting the mean value equal to unity the origin becomes a point of zero error while the tolerance and coverage are altered to some extent. The choice of the mean value thus depends upon the detailed requirements of a specific design.

IX) in connection with the design of filter networks. There we deal with functions of the form given by equation (866), page 395, where the approximation range is defined by $-x_1 \leq x \leq x_1$ (x_1 being equal to unity). If x_1 is replaced by x_{n+1}, where n equals the number of coincident poles and zeros of Z_a', then a plot of the irrational factor $\sqrt{1-x^2/x_1^2}$ becomes very similar in form to the normalized ratio discussed here for the case where all the poles and zeros of Z_a' are coincident with those of the tangent function.[1] Such a plot is shown by the dotted curve of Fig. 97. Thus the methods of parameter determination given there may serve as first approximations in the solution to our present problem.

The highest frequency for which the artificial line is designed is evidently given by

$$\omega_h = k\omega_{n+1} = \frac{\pi k(n+1)}{l\sqrt{LC}}$$

which may also be solved for n when the other factors are prescribed. The frequency range is thus primarily determined by the number of *coincident* zeros and poles, while the tolerance and coverage are fixed by the additional *non-coincident* ones.

The manner in which the effect of dissipation may properly be taken into account becomes clear if we recall the discussion given toward the end of section 8 of the preceding chapter, and particularly in the footnote on page 245. These considerations suggest that the uniform dissipation of the line is accurately accounted for by adding series resistances to all the coils and parallel conductances to all the condensers in amounts such that all the resistance-to-inductance ratios equal R/L and all the conductance-to-capacitance ratios equal G/C where these are the ratios for the line which is to be simulated.

The correctness of this procedure is seen from the following argument. If we set up the sinusoidal steady-state mesh charge equations for such a uniformly dissipative network, the determinant D^* of this set of equations has elements of the form

$$b_{ik}^* = L_{ik}(j\omega + 2\alpha)(j\omega + 2\beta) + S_{ik}, \tag{621}$$

where

$$\alpha = \frac{R}{2L}; \; \beta = \frac{G}{2C} \tag{621a}$$

are the uniform ratios. The *charge impedances* (ratio of complex volt-

[1] This applies to cases where n is odd. When n is even the plot is similar to the reciprocal of the irrational factor as may be seen from Fig. 98.

age amplitude to complex mesh-charge amplitude) of this network have the form

$$Z_{ik}^* = \frac{D^*}{B_{ik}^*} \tag{621b}$$

where B_{ik}^* is a minor of D^*. Since a complex mesh current equals a complex mesh charge multiplied by $(j\omega + 2\beta)$, a *current impedance* is given by

$$Z_{ik} = \frac{Z_{ik}^*}{j\omega + 2\beta}. \tag{621c}$$

For the corresponding non-dissipative case the same relations apply with $\alpha = \beta = 0$.

This situation makes it clear that the charge impedance Z_{ik}^* for a uniformly dissipative network is obtained from that for the corresponding non-dissipative network by replacing $j\omega$ in the non-dissipative function by $\sqrt{(j\omega + 2\alpha)\,(j\omega + 2\beta)}$. Hence the current impedance Z_{ik} for a uniformly dissipative network is obtained from the corresponding non-dissipative impedance function by replacing $j\omega$ by $\sqrt{(j\omega + R/L)\,(j\omega + G/C)}$ and then multiplying by the factor

$$\sqrt{\frac{j\omega + 2\alpha}{j\omega + 2\beta}} = \sqrt{\frac{j\omega + R/L}{j\omega + G/C}}. \tag{621d}$$

Since a propagation function depends only upon the ratio of two impedances, the dissipative is obtained from the corresponding non-dissipative simply by replacing $j\omega$ by $\sqrt{(j\omega + R/L)(j\omega + G/C)}$.

The fact that this same procedure applies to the long-line propagation and characteristic impedance functions shows that the uniformly dissipative artificial line accurately simulates the dissipative properties of the actual line.

It should be noted in this connection that if the lattice reactances satisfy the relation $Z_aZ_b = Z_0^2$ in the non-dissipative case (where $Z_0 = \sqrt{L/C}$), then in the uniformly dissipative lattice the component impedances satisfy the same relation for Z_0 equal to its dissipative value. That is, the uniformly dissipative lattice has exactly the same characteristic impedance as the dissipative line, the whole error being in the propagation function as given by the non-dissipative design.

It is interesting to note that this latter method of artificial line design is also a method for designing linear phase-shift networks, i.e., networks having a constant characteristic impedance, zero attenuation, and a linear phase characteristic versus frequency. This is clear from the fact that a non-dissipative line fulfils these specifications. Net-

works of this sort may be used to introduce a specified amount of time delay, the latter being equal to the slope of the phase characteristic and hence equal to $l \cdot \sqrt{LC}$ seconds.

In this connection, as well as for the design of artificial lines on the lattice basis, it is important to note the following relationship between the error ϵ in the reactance Z_a' and the corresponding error in the resulting phase function of the lumped structure. Writing the first relation (615) as

$$Z_a = jZ_0 \tan \frac{\alpha_2 l}{2},$$

we get by forming the differential

$$dZ_a = j\frac{Z_0 l}{2} \cdot \frac{d\alpha_2}{\cos^2 \dfrac{\alpha_2 l}{2}}.$$

Since for small errors we can write $Z_a' = Z_a + dZ_a$, we have

$$\epsilon = \frac{dZ_a}{Z_a} = \frac{l d\alpha_2}{\sin \alpha_2 l},$$

so that the resulting deviation in the phase function of the lattice becomes

$$d\alpha_2 = \frac{\epsilon \sin \alpha_2 l}{l}.$$

Since α_2 is substantially a linear function of frequency, we see that, for a constant error ϵ, the deviation in the phase function becomes a sinusoidal function of frequency. This is a novel result compared with that obtained with the uniform ladder structure where, according to (602) and (603), the deviation is a continuously increasing function of frequency.

This fact makes it necessary to define the error in the phase function of the lattice on a different basis. Since the variation in the *slope* of the phase function determines the extent of phase distortion in a network (as will be discussed in detail in Chapter XI), it is not sufficient to regard the maximum decimal deviation in α_2 as the criterion of quality when that deviation is oscillatory. Instead, we must define the decimal error in the resulting phase function in terms of its slope as compared with that of the desired function, i.e., as

$$\delta_{\alpha_2} = \frac{\dfrac{d\alpha_2'}{d\omega} - \dfrac{d\alpha_2}{d\omega}}{\dfrac{d\alpha_2}{d\omega}},$$

where $\alpha_2' = \alpha_2 + d\alpha_2$ is the resulting phase function for small deviations.

Since $\alpha_2' = \alpha_2 + d\alpha_2$, we have according to the above

$$\frac{d\alpha_2'}{d\omega} = \frac{d\alpha_2}{d\omega}(1 + \epsilon \cos \alpha_2 l),$$

and the correct measure of quality becomes

$$\delta_{\alpha_2} = \epsilon \cos \alpha_2 l,$$

the maximum value of which is ϵ, which is the maximum error in the reactance Z_a'!

3. Lump-loaded lines. In Chapter III we pointed out the desirability of increasing the distributed inductance of lines or cables for the purpose of decreasing the attenuation as well as the amount of distortion over a given frequency range. Although this may be accomplished by wrapping the conductors with a material of high permeability, such a procedure is both difficult and expensive, and is at present employed only for submarine circuits where there is no practical alternative. On land circuits the situation may be met in another way, namely, by inserting lumped inductances at intervals. This process becomes more nearly equivalent to the distributed circuit the shorter the lumped intervals are made.

The reason why a very lumpy distribution of inductance will not be satisfactory should be clear from what has just been said regarding artificial lines of the ladder type. These are only approximately equivalent to the smooth line over a limited frequency range, the discrepancy becoming larger at higher frequencies and finally exceeding a prescribed maximum tolerance. Just so, a lumpy loaded line will be only approximately equivalent to the smoothly loaded line; and the discrepancy will increase with frequency until it exceeds a set upper limit.

In determining the amount of loading necessary in a certain case, we supposed that certain upper limits had been set for the tolerances in the propagation and characteristic impedance functions. In order to be consistent we should now continue the problem from the same point of view and attempt to find the largest spacing for the loading coils for which the same maximum tolerances will not be exceeded below an upper frequency limit.

Suppose we denote the amount of additional inductance (loading coil inductance) per mile of line by L_l, and the required number of loading coils (loading sections) per mile by n. The distributed impedance and admittance per mile of original line are denoted by z and y as before. Fig. 99 then illustrates what is known as one **loading section.**

This is taken as that portion of the original line included between two successive loading coils plus one-half of a loading coil at each end. The loaded line may evidently be thought of as a cascade of identical sections of this type. We might just as well have defined a loading section as that portion of loaded line included between two successive midpoints of smooth line. This would lead to an analysis on the basis of Π-sections whereas the basis adopted in Fig. 99 makes use of the T-section.

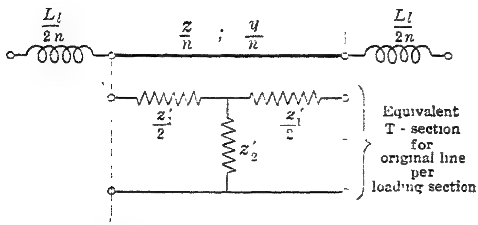


FIG. 99.—Schematic of a mid-coil loading section.

Our method of procedure is to find the equivalent lumped T-section for the loading section, and then write down the expressions for the propagation and characteristic impedance functions for this equivalent lumped loading section. A comparison of these functions with those for the uniformly loaded line will enable us to study the discrepancies as functions of the number of loading sections per mile.

The first step in this procedure is to determine the lumped equivalent to that portion of original smooth line included between two loading coils. Using the relations (594) and the notation of Fig. 99 we have

$$\left.\begin{aligned} z_1' &= 2Z_0 \tanh \frac{\alpha}{2n} \\ z_2' &= \frac{Z_0}{\sinh \frac{\alpha}{n}} \end{aligned}\right\}, \tag{622}$$

where $\alpha = \sqrt{yz}$ is the original propagation function of the line.

Since we wish to add to this lumped section the inductances of the loading-coil halves at either end, it will be more convenient to do away with the hyperbolic functions. Recognizing that α/n must be a small quantity, we can accomplish this by series expansions. If we do this and drop all terms above the square in α/n we get

$$\left.\begin{aligned} z_1' &\cong \frac{z}{n}\left(1 - \frac{1}{12}\frac{\alpha^2}{n^2}\right) \\ z_2' &\cong \frac{n}{y}\left(1 - \frac{1}{6}\frac{\alpha^2}{n^2}\right) \end{aligned}\right\}. \tag{622a}$$

Denoting the series and shunt impedances of the T-section equivalent to the loading section of Fig. 99 by z_1 and z_2, respectively, we have

$$\left.\begin{aligned} z_1 = z_1' + j\omega\frac{L_l}{n} &= \frac{z'}{n}\left\{1 - \frac{1}{12}\left(\frac{z}{z'}\right)\left(\frac{\alpha}{n}\right)^2\right\} \\ z_2 = z_2' &= \frac{n}{y}\left\{1 - \frac{1}{6}\left(\frac{\alpha}{n}\right)^2\right\} \end{aligned}\right\}, \tag{623}$$

where we have let

$$z' = z + j\omega L_l \tag{624}$$

represent the series impedance per mile of uniformly loaded line.

With the impedances (623) we must now find the propagation and characteristic impedance functions for the equivalent loading section. For this purpose we use the expressions (597) and (598) which we repeat below

$$Z_T = \sqrt{z_1z_2}\sqrt{1 + \frac{z_1}{4z_2}}, \tag{597}$$

$$\gamma = 2\sinh^{-1}\sqrt{\frac{z_1}{4z_2}}. \tag{598}$$

From (623) we find by means of expansions

$$\frac{z_1}{4z_2} = \frac{1}{4}\left(\frac{\alpha'}{n}\right)^2\left\{1 + \frac{1}{6}\left(\frac{\alpha}{n}\right)^2\left(1 - \frac{z}{2z'}\right) + \cdots\right\}, \tag{625}$$

$$\sqrt{1 + \frac{z_1}{4z_2}} = 1 + \frac{1}{8}\left(\frac{\alpha'}{n}\right)^2 + \cdots, \tag{625a}$$

and

$$\sqrt{z_1z_2} = Z_0'\left\{1 - \frac{1}{12}\left(\frac{\alpha}{n}\right)^2\left(1 + \frac{z}{2z'}\right) + \cdots\right\}, \tag{625b}$$

where

$$\alpha' = \sqrt{yz'} \quad\text{and}\quad Z_0' = \sqrt{\frac{z'}{y}} \tag{626}$$

are the propagation and characteristic impedance functions for the uniformly loaded line.

Using these results in (597) we find for the characteristic impedance of the lump-loaded line

$$Z_T = Z_0'\left\{1 + \frac{\alpha'^2 - \alpha^2}{8n^2}\left(1 + \frac{z}{3z'}\right) + \cdots\right\}. \tag{627}$$

For the propagation function per mile of the lump-loaded line we find by using (625) in an expansion of (598)

$$n\gamma = \alpha'\left\{1 - \frac{\alpha'^2 - \alpha^2}{24n^2}\left(1 - \frac{z}{z'}\right) + \cdots\right\}. \tag{628}$$

The results (627) and (628) show that the lump-loaded line is equivalent to the uniformly loaded line to within an approximate error in the characteristic impedance function of

$$\delta_z = \frac{\alpha'^2 - \alpha^2}{8n^2}\left(1 + \frac{z}{3z'}\right), \tag{629}$$

and in the propagation function of

$$\delta_\gamma = -\frac{\alpha'^2 - \alpha^2}{24n^2}\left(1 - \frac{z}{z'}\right). \tag{630}$$

In terms of the original line parameters and the additional loading inductance these are

$$\delta_z = \frac{j\omega L_l(G + jC\omega)}{8n^2}\left(1 + \frac{1}{3}\cdot\frac{R + jL\omega}{R + j(L + L_l)\omega}\right), \tag{629a}$$

and

$$\delta_\gamma = -\frac{j\omega L_l(G + jC\omega)}{24n^2}\left(1 - \frac{R + jL\omega}{R + j(L + L_l)\omega}\right). \tag{630a}$$

These errors are evidently zero at zero frequency[1] and increase with frequency. They will, therefore, have their maximum values at the upper essential frequency. At this frequency it will usually be possible to neglect R and G. Denoting the upper essential frequency by ω_2, we then have

$$\delta_z = -\frac{L_l C\omega_2^2}{8n^2}\left(1 + \frac{L}{3(L + L_l)}\right), \tag{629b}$$

and

$$\delta_\gamma = \frac{L_l C\omega_2^2}{24n^2}\left(1 - \frac{L}{L + L_l}\right). \tag{630b}$$

Usually $L_l \gg L$ so that the parentheses in these expressions become unity. Without regard to the sign of δ_z we then have for the number of loading sections per mile

$$n = \frac{\omega_2\sqrt{L_l C}}{2\sqrt{2\delta_z}} \quad \text{or} \quad n = \frac{\omega_2\sqrt{L_l C}}{2\sqrt{6\delta_\gamma}}. \tag{631}$$

[1] This is due here to the fact that we have neglected the resistance of the loading coils, which is usually a justified procedure. The student may derive corresponding relations taking this resistance into account and compare results for typical practical parameter values.

These results show that δ_z is about three times the error in the propagation function. From what was said in section 8 of Chapter III we recognize that the error in the propagation function will usually be the governing factor. Whether this is true or not will decide, of course, which of the expressions (631) is to be used for the determination of n.

As an illustrative example let us refer again to the problem discussed in section 5, Chapter III (page 98), where the necessary R/L ratio to make $\delta_\alpha = 0.05$ with $\omega_1 = 2\pi \cdot 100$ was calculated for the open line data on page 90. The results showed that the inductance should be 3.78 times the old value. This makes the loading coil inductance

$$L_l = 2.78\ L = 0.0111 \text{ henry per mile.}$$

The approximations involved in dropping the parentheses in (629b) and (630b) are not justified. Taking the upper frequency limit as

$$\omega_2 = 2\pi \cdot 5000,$$

and using (630b) with $\delta_\gamma = 0.05$, i.e., the same maximum error as was used for the determination of the amount of loading, we find

$$n = 0.232,$$

which means that the specifications will be met by placing a coil with 0.048 henry every 4.3 miles

4. Cut-off frequency of lump-loaded lines. Lump-loaded lines as well as artificial lines of the ladder type possess an additional important property which becomes evident at frequencies considerably higher than the upper limit in the above analysis. At these frequencies we may neglect R and G without appreciable error. If we assume that $L_l \gg L$, then we have for the impedances z_1 and z_2 of the loading section according to (623)

$$z_1 = \frac{jL_l\omega}{n};\ z_2 = \frac{n}{jC\omega},$$

so that

$$\sqrt{\frac{z_1}{4z_2}} = \frac{j\omega\sqrt{L_lC}}{2n}.$$

Recognizing that the propagation function of the loading section is in general complex, although in this case we expect it to be purely imaginary on account of having neglected dissipation, we write

$$\gamma = \gamma_1 + j\gamma_2,$$

and get by means of (598)

$$\frac{j\omega\sqrt{L_lC}}{2n} = \sinh\frac{\gamma_1}{2}\cos\frac{\gamma_2}{2} + j\cosh\frac{\gamma_1}{2}\sin\frac{\gamma_2}{2},$$

where the formula for the hyperbolic sign of the sum of two angles is used. Equating reals and imaginaries gives

$$\left.\begin{aligned} \sinh\frac{\gamma_1}{2}\cos\frac{\gamma_2}{2} &= 0 \\ \cosh\frac{\gamma_1}{2}\sin\frac{\gamma_2}{2} &= \frac{\omega\sqrt{L_lC}}{2n} \end{aligned}\right\}. \tag{632}$$

These equations are simultaneously satisfied by either

$$\gamma_1 = 0;\ \gamma_2 = 2\sin^{-1}\frac{\omega\sqrt{L_lC}}{2n}, \tag{633}$$

or

$$\gamma_2 = \nu\pi;\ \gamma_1 = 2\cosh^{-1}\frac{\omega\sqrt{L_lC}}{2n}\cdot(-1)^{\frac{\nu-1}{2}}, \tag{633a}$$

where ν is an *odd* integer. This result shows that the loading section does not possess zero attenuation for all frequencies even though dissipation is neglected. By (633) we see that $\gamma_1 = 0$ only so long as the equation for γ_2 can be satisfied, which is for frequencies within the range

$$-1 \leq \frac{\omega\sqrt{L_lC}}{2n} \leq 1. \tag{634}$$

This fixes a definite upper frequency limit, namely

$$\omega_c = \frac{2n}{\sqrt{L_lC}} \tag{635}$$

above which the attenuation will rise according to the second equation (633a). The frequency range $0 \leq \omega \leq \omega_c$ is called the *transmission range,* while that defined by $\omega_c \leq \omega < \infty$ is referred to as the *attenuation range* for the loaded line. ω_c is called the **critical** or also the **cut-off frequency.** These matters have already been mentioned with regard to the general four-terminal network in section 9 of Chapter IV.

A similar situation holds for the artificial line in the form of a ladder network.

It is interesting to evaluate the errors δ_z and δ_γ at this cut-off frequency. Using (635) in (633) we have for the propagation function per mile of the lump-loaded line at $\omega = \omega_c$

$$n\gamma_c = jn\pi,$$

while by (626) we have for the uniformly loaded line at this frequency (neglecting losses)

$$\alpha_c' = j2n.$$

The error is determined from

$$n\gamma_c = \alpha_c'(1 + \delta_{\gamma c}),$$

which gives

$$\delta_{\gamma c} = \frac{\pi}{2} - 1 = 57\%.$$

For the evaluation of the error in the characteristic impedance at ω_c we note that at this frequency

$$\frac{z_1}{4z_2} = -1,$$

so that $Z_T = 0$, and hence from

$$Z_T = 0 = Z_0'(1 + \delta_{zc})$$

we get

$$\delta_{zc} = -100\%.$$

In practice the relation (635) is sometimes used for the determination of the number of loading coils per mile with ω_c set equal to the upper essential frequency. Since this assumes considerable maximum errors in the propagation and characteristic impedance functions, namely 57 per cent and 100 per cent respectively, n by this method comes out smaller. For the above numerical case we get

$$n = 5000\pi\sqrt{.0111 \cdot 0.008} \cdot 10^{-3} = 0.148,$$

which requires loading coils of 0.075 henry at intervals of 6.76 miles.

PROBLEMS TO CHAPTER VII

7-1. For the open line whose data are specified in Problem 3-1 it is desired to design an artificial line with a maximum departure in the propagation function of 2 per cent over a frequency range from 50 to 5000 cycles per second. Carry out the design assuming a cascade of identical T-sections.

7-2. Repeat the design of Problem 7-1 for the cable data specified in Problem 3-2.

7-3. Carry out the design of the artificial line specified in Problem 7-1 by assuming a lattice structure and utilizing either of the expansions (610) or (611). In this method of procedure the error may be determined by considering first the corresponding non-dissipative line and determining the maximum departure from linearity in the phase function according to the method discussed in the latter part of section 2. Note that the method of synthesizing z_a according to the expansions (610) and (611) then becomes identical with the method of Foster's theorem. Note also the bearing which the result of Problem 6-7 has upon the corresponding dissipative case.

7-4. Continue the design of the artificial line according to Problem 7-3 by introducing non-coincident critical frequencies instead of having only coincident ones

according to the expansions (610) and (611). Do this for the non-dissipative case and note the increased economy which results. Properly introduce the effect of dissipation and compare the final design with that found according to the procedure followed in Problem 7–1.

7–5. Transform the lattice design of Problem 7–4: (*a*) into an equivalent bridged-T and (*b*) into an equivalent cascade of simpler lattice structures according to the method discussed in the latter part of section 2, Chapter XII, and note whether any increase or decrease in the total number of network elements results.

7–6. Repeat the artificial line design of Problem 7–4 for the cable data referred to in Problem 7–2, and compare with the design obtained for that problem.

7–7. Design a non-dissipative linear phase-shift network on the lattice basis to have a basic delay of 1 millisecond and a maximum departure from this delay of 4 per cent over a frequency range from 0 to 8000 cycles per second.

7–8. Referring to Problem 3–7, determine the proper size and spacing of the loading coils so that the specified tolerance is not exceeded at any point in the essential frequency range. Find the theoretical cut-off frequency of the corresponding non-dissipative facility.

7–9. Reconsider the statement of Problem 3–9, but do *not* assume the effect of the loading to be the same as for a corresponding uniformly loaded circuit.

CHAPTER VIII

THE NON-DISSIPATIVE UNIFORM LADDER STRUCTURE

1. Formulation of the problem. In this chapter we shall study the uniform ladder structure with non-dissipative component impedances. We shall consider its steady-state behavior when excited at one end by means of a harmonic generator with internal impedance Z_S, while the far end is terminated in an impedance Z_R. We picture the network as a whole very much like that of the uniform line with sending- and receiving-end impedances.

There is a distinct difference between the long-line picture and the present one, which presents itself at the outset. The long line is strictly uniform in the sense that its terminals have the same appearance no matter where the structure is broken off. This is no longer true of the ladder network of which Fig. 100 shows an internal section. The terminals may present an infinite variety of forms depending upon the points at which the structure is broken off. We may, for example, terminate with any portion or all of the last series impedance attached to the structure. Similarly we may terminate with any fraction of the shunt admittance in the last "rung" of the ladder.

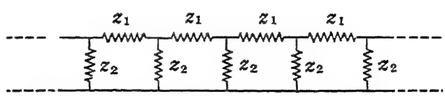


FIG. 100.—Schematic of a uniform recurrent ladder network.

The terminal conditions and hence the behavior of the complete network will depend upon the exact manner in which the structure is terminated.

Out of the infinite number of possible terminations which present themselves, two typical ones are of primary importance. These we shall discuss in some detail, and leave the investigation of **fractional terminations,** as the intermediate ones are called, for consideration later.

Suppose we imagine two successive series impedances bisected and the intervening portion of the network removed. This portion is evidently a symmetrical T-structure. For fairly obvious reasons it is also referred to as a **mid-series section.** If the ladder is broken off with half a series impedance attached, we speak of this end as being *mid-series terminated.* If the structure is terminated in this way at both

ends, we may look upon the ladder as a cascade connection of symmetrical T-sections or mid-series sections.

The other type of termination to which we refer is the so-called **mid-shunt** termination which is obtained by breaking off the ladder so as to leave half of (twice) the last shunt admittance (impedance) attached. By passing cleavage planes through the centers of two successive shunt admittances and removing the intervening portion, we evidently get a symmetrical Π-structure, which is also referred to as a **mid-shunt section.** A ladder which is mid-shunt terminated at both ends may be looked upon as a cascade connection of symmetrical Π-sections or mid-shunt sections.

The behavior of the ladder network terminated either in mid-series or mid-shunt at both ends, and working out of and into impedances Z_S and Z_R, is, therefore, given by the discussion in section 8 of Chapter IV where the cascade connection of arbitrary four-terminal networks is considered. In the present case there is no distinction between the iterative and image bases because the component networks are symmetrical. The solution for either case may, therefore, be used.

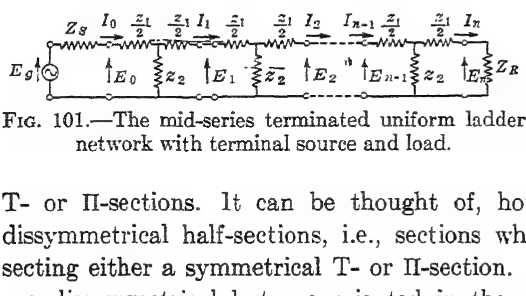


FIG. 101.—The mid-series terminated uniform ladder network with terminal source and load.

When the ladder is terminated in mid-series on one end and in mid-shunt on the other, it can no longer be thought of simply as a cascade of either T- or Π-sections. It can be thought of, however, as a cascade of dissymmetrical half-sections, i.e., sections which are obtained by bisecting either a symmetrical T- or Π-section. Since such half-sections are dissymmetrical but are oriented in the resulting ladder so that adjacent ends are similar, we have a cascade network built up on the image basis, and hence the forms (400), (401), (402), and (403) must be applied.

In the present chapter we shall give specific attention only to the case with mid-series terminations at both ends. On the basis of the general treatment in section 8 of Chapter IV, the student may carry out the details of the remaining cases as practice problems.

2. The mid-series terminated structure. Fig. 101 illustrates the physical situation. The notation regarding junctions is the same as that previously used for the general four-terminal networks. The solutions (376), page 165, for the iterative case may be applied here where no distinction need be made between the characteristic impedances Z_{01}

and Z_{02}. For their common value we shall write Z_T because the T-section is involved. We, therefore, have

$$\left.\begin{aligned} E_k &= \frac{Z_T E_g(e^{(n-k)\gamma} + r_R e^{-(n-k)\gamma})}{(Z_S + Z_T)(e^{n\gamma} - r_S r_R e^{-n\gamma})} \\ I_k &= \frac{E_g(e^{(n-k)\gamma} - r_R e^{-(n-k)\gamma})}{(Z_S + Z_T)(e^{n\gamma} - r_S r_R e^{-n\gamma})} \end{aligned}\right\}, \tag{636}$$

where for the propagation and characteristic impedance functions we have by (424) and (425a), page 178,

$$\left.\begin{aligned} \gamma &= 2 \sinh^{-1}\sqrt{\frac{z_1}{4z_2}} \\ Z_T &= \sqrt{z_1 z_2}\sqrt{1 + \frac{z_1}{4z_2}} \end{aligned}\right\}, \tag{637}$$

and the reflection coefficients are given by

$$r_S = \frac{Z_S - Z_T}{Z_S + Z_T}; \quad r_R = \frac{Z_R - Z_T}{Z_R + Z_T}. \tag{638}$$

From these the various voltage, current, and volt-ampere ratios may be formed just as this was done for the long line in section 17 of Chapter II. We shall discuss in detail only one such ratio, namely, the ratio of the output voltage E_n to the generator voltage E_g. By analogy to the long-line problem we have for the present case with the use of (128), page 72,

$$\frac{E_n}{E_g} = \frac{Z_R}{Z_S + Z_R} \cdot e^{-n\gamma} \cdot (1 - r_S r_R) \cdot (1 - r_S r_R e^{-2n\gamma})^{-1}. \tag{639}$$

This ratio must be discussed separately for the transmission and attenuation ranges. We shall begin with the former.

In the non-dissipative structure the transmission range is that range for which the attenuation function is zero.
That is

$$\gamma = 0 + j\gamma_2.$$

Using this in the first equation (637) we get

$$\frac{z_1}{4z_2} = -\sin^2\frac{\gamma_2}{2}, \tag{640}$$

from which we see that

$$-1 \leq \frac{z_1}{4z_2} \leq 0 \tag{641}$$

defines the transmission region; i.e., the frequency range for which (641) is satisfied becomes a transmission region in the sense that $\gamma_1 = 0$

for all frequencies within this range. The object of the present discussion is to give a more physical interpretation to this range.

Let us consider the case for which the terminal impedances Z_S and Z_R are pure resistances. This is of primary interest in practice. Also let us limit our discussion for the moment to the magnitude of the ratio (639). Then we have for the transmission region

$$\left|\frac{E_n}{E_g}\right| = \frac{Z_R}{Z_S + Z_R} \cdot (1 - r_S r_R) \cdot \left|1 - r_S r_R e^{-2jn\gamma_2}\right|^{-1}. \qquad (639a)$$

Here it is important to note that, since z_1 and z_2 are pure reactances, the condition (641) shows that the characteristic impedance according to (637) is purely real in any transmission range. Hence, with pure resistance terminations, the reflection coefficients become real numbers throughout this range. From the nature of reactance functions we note that Z_T will vary uniformly[1] with frequency and will pass continuously from the value $\sqrt{z_1 z_2}$ to zero as $z_1/4z_2$ passes from zero to minus one. The corresponding variations in r_S and r_R will also be uniform, and will pass from certain initial values to plus one throughout the same range.

This rather slow variation with frequency has only a secondary effect upon the function (639a). Its principal variation in the case of a sufficient number of sections is due to the variation of γ_2. As the parameter $z_1/4z_2$ passes from zero to minus one, γ_2, according to (640), passes from zero to $\pm\,\pi$. The exponential factor

$$e^{-2jn\gamma_2},$$

therefore, makes n revolutions throughout this range. Thus the quantity

$$|1 - r_S r_R e^{-2jn\gamma_2}|$$

is seen to pass through n maximum and n minimum values for such a range. This in turn causes the voltage ratio to fluctuate between the same number of maxima and minima. Since the variation of $r_S r_R$ is slow, the frequencies at which these maxima and minima occur are very nearly determined by the variation in γ_2 alone. This we shall assume to be sufficiently close, and hence obtain the result that extremal values of the function (639a) occur for

$$e^{-2jn\gamma_2} = \pm\,1. \qquad (642)$$

In particular we have

$$e^{-2jn\gamma_2} = +\,1 \text{ for } \gamma_2 = \frac{\nu\pi}{2n} \qquad (642a)$$

[1] This assumes the general case for which z_1 and z_2 have no coincident zeros or poles. Then the variation in Z_T cannot be oscillatory in any transmission range.

when ν is an *even* integer, and

$$e^{-2jn\gamma_2} = -1 \text{ for } \gamma_2 = \frac{\nu\pi}{2n} \tag{642b}$$

when ν is an *odd* integer. From (639*a*) the corresponding extremal values of the voltage ratio are

$$\left|\frac{E_n}{E_g}\right|_\nu = \frac{Z_R}{Z_S + Z_R}; \ (\nu \text{ even}), \tag{643}$$

and

$$\left|\frac{E_n}{E_g}\right|_\nu = \frac{Z_R Z_T}{Z_S Z_R + Z_T^2}; \ (\nu \text{ odd}), \tag{643a}$$

where the relations (638) are made use of. The frequencies at which these extrema occur are found from (640) together with (642*a*) and (642*b*). They are determined by the equation

$$\frac{z_1}{4z_2} + \sin^2\left(\frac{\nu\pi}{4n}\right) = 0. \tag{644}$$

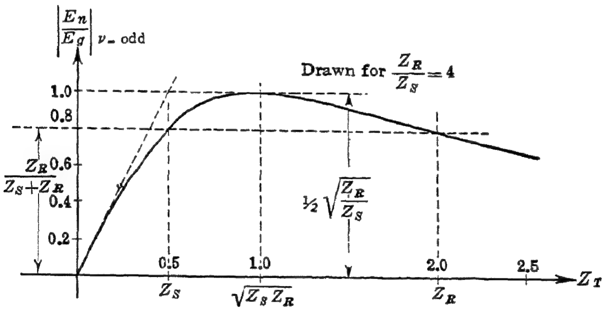


FIG. 102.—Response ratio for natural frequencies of odd order according to (643*a*).

A good idea of the behavior of the voltage ratio in the transmission range may be had by drawing the loci (643) and (643*a*), and marking points on these corresponding to the frequencies determined from (644). A continuous curve alternately joining the points on the two loci then gives a sufficiently good picture of the resulting behavior versus frequency.

Some remarks are in order regarding the loci of extremal values. Since the terminal impedances are assumed constant, the locus (643) is a horizontal line, and needs no further comment. The locus (643*a*), however, depends upon Z_T and hence varies with frequency. Let us study this locus first as a function of Z_T. When, for a specific case, Z_T is given as a function of frequency, the proper function of frequency for the locus may then easily be obtained either graphically or analytically.

Fig. 102 shows a plot of (643a) as a function of Z_T. We see that the function has a maximum where Z_T equals the geometric mean between the terminal resistances, and that the value of this maximum equals one-half the square root of the ratio of Z_R to Z_S. The horizontal dotted line represents the locus (643). The locus for ν-odd, therefore, cuts that for ν-even at two points which are determined by equating the expressions for the two loci, thus

$$\frac{Z_R}{Z_S + Z_R} = \frac{Z_R Z_T}{Z_S Z_R + Z_T^2}.$$

The fact that this quadratic in Z_T has real roots for all possible combinations of Z_S and Z_R shows that the intersections always exist. These roots are easily found to be

$$Z_T = Z_S \quad \text{and} \quad Z_T = Z_R.$$

An exception occurs only in the special case $Z_S = Z_R$. Then the two intersections merge with the maximum; i.e., the loci become tangent at the point $Z_T = Z_S = Z_R$.

The locus for ν-even is thus not the locus of maxima and that for ν-odd the locus of minima. The rôles of maxima and minima are in general shared by the two loci.

Before we may proceed with an illustrative example, it is necessary to call attention to an additional point. This has to do with the locus (643), which is that for even integer values of ν. If we examine (644), we note that the boundary of the transmission region at which

$$\frac{z_1}{4z_2} = -1$$

corresponds to ν-even regardless of whether n is even or odd. This boundary frequency, therefore, should belong to the locus (643). However, since at this boundary $\gamma_2 = \pm \pi$ and $Z_T = 0$ so that $r_S = r_R = 1$, we note that the ratio (639a) becomes indeterminate, and hence its value is not necessarily given by (643). This must be investigated.

For the evaluation of this indeterminate it is convenient to throw (639a) into the alternative form

$$\left|\frac{E_n}{E_g}\right| = \frac{Z_R}{\left|(Z_S + Z_R)\cos n\gamma_2 + j\left(\frac{Z_S Z_R}{Z_T} + Z_T\right)\sin n\gamma_2\right|} \tag{645}$$

which is obtained with the use of (638) and writing the sine and cosine equivalent for the exponential. Making use of the fact that for the symmetrical non-dissipative T-section

$$Z_T = jz_2 \sin\gamma_2, \tag{646}$$

as may be seen from the second equation (426), page 178, the form (645) becomes

$$\left|\frac{E_n}{E_g}\right| = \frac{Z_R}{\left|(Z_S + Z_R)\cos n\gamma_2 + \left(\frac{Z_S Z_R}{z_2}\right)\frac{\sin n\gamma_2}{\sin\gamma_2} - z_2 \sin n\gamma_2 \sin\gamma_2\right|}. \tag{645a}$$

The desired value is now obtained by evaluating this for the limit $\gamma_2 \to \pm\pi$. If we denote this limiting value by the subscript c, and write

$$z_2 = j|z_2|,$$

the magnitude of the limit becomes

$$\left|\frac{E_n}{E_g}\right|_c = \frac{Z_R}{Z_S + Z_R}\left\{1 + \frac{n^2 Z_S^2 Z_R^2}{|z_2|_c^2 (Z_S + Z_R)^2}\right\}^{-\frac{1}{2}}. \tag{647}$$

This is the value of the locus for ν-even at the boundary where

$$\frac{z_1}{4z_2} = -1.$$

Comparing this result with the locus (643), we see that the boundary value is in general smaller, and becomes smaller as n increases. Note that if $Z_S = 0$, (647) becomes equal to (643).

For a very common type of filter network, which we shall discuss in detail in the next chapter, the component reactances z_1 and z_2 are inverse networks of impedance product R^2 where $R = Z_S = Z_R$. At the boundary in question, therefore, $|z_2|_c^2 = R^2/4$, so that for this type

$$\left|\frac{E_n}{E_g}\right|_c = \frac{1}{2\sqrt{1 + n^2}}. \tag{647a}$$

3. Physical interpretation. The results obtained so far for the behavior of the non-dissipative uniform recurrent ladder structure may be given a somewhat enlightening interpretation in terms of more familiar network principles. Such an interpretation seems in order at this time for several reasons. We saw in section 9 of Chapter IV that the general non-dissipative four-terminal network can, under certain circumstances and for definite ranges, exhibit transmission properties, i.e., possess zero attenuation. We met this phenomenon again in the discussion of lump-loaded lines. The condition for the existence of such a transmission range, like the relation (641) in the present instance, is merely a mathematical formality and offers no substantial inkling as to the probable physical reason for such behavior. This situation may now be made somewhat clearer.

Namely, the fact that the network response within the transmission range shows a tendency to fluctuate between maxima and minima leads us to interpret this behavior as an exhibition of resonance phenomena. In this sense the transmission range of the structure immediately assumes the very physical aspect of being that region within which the natural frequencies lie. In other words, a transmission range, or, in filter terminology, the pass band, *is a resonance region for the network.*

We met examples of this much earlier in our study of networks. At the close of Chapter III and again in Chapter VIII (both in Vol. I), we studied the selective properties of one or more tuned coupled circuits and found that they answered very well as filters for the selection of a band of frequencies. There, this selective property could easily be seen to be due to a resonance phenomenon. In meeting this property again in lump-loaded lines, in artificial lines, and in the present ladder structure, the mathematical form of expression—which leans heavily toward that found for the long-line behavior—tends to obscure the more familiar lumped-constant network behavior, but the physical explanation of the selective properties in terms of resonance phenomena is none the less real, and should certainly not be lost sight of.

A filter or selective system differs essentially from a perfectly arbitrary lumped network only in one particular. The arbitrary network has its natural frequencies distributed promiscuously along the frequency axis, whereas the selective system is so constituted as to have them arranged in groups which become the regions of exceptional response, or, as we then prefer to call them, the transmission regions or pass bands. Broadly speaking, the problem of filter design is to determine the structure of a network in such a way that its natural frequencies will be wholly confined to lie within prescribed frequency limits.

This, at least, is a quite obvious requirement for any selective system. The second no less important requirement is that the exceptional response of the network within any resonance region shall be as uniform as possible. This requirement is the one which is more difficult to meet.

For the ladder structure discussed here, the relation (644) defines not only the natural frequencies at which resonance occurs, but also those intervening frequencies at which the response tends to subside to less exceptional values. With pure reactive component impedances we note that a resonance region will lie in the vicinity of each zero of z_1[1], because there the parameter $z_1/4z_2$ will evidently pass through the values -1 to 0. The number of critical frequencies which (644) will thus deter-

[1] Provided z_2 does not have a zero at the same point.

mine in this vicinity depends upon n, i.e., upon the number of T-sections in cascade.

The second requirement which has to do with the uniformity of the resonance response must be investigated in terms of the loci of extremal values, namely the loci (643) and (643a). The ideal would be to have these loci coincide in the same horizontal line, which means that (643a) would have to be made to coincide with (643). This requires that either

$$Z_{\mathrm{T}} = Z_S$$

or

$$Z_{\mathrm{T}} = Z_R.$$

Since the terminal impedances are constant, this means that the characteristic impedance must be constant throughout the transmission range. Although this condition cannot be fully attained, it may be approximated to any desired extent by the choice of suitable component reactances. The problem is complicated by the fact that while this adjustment is sought, the behavior in the attenuation region must not be lost sight of. The desired end may be reached in a number of ways. This topic will receive our attention throughout the following two chapters. At present we wish to point out, first, to what extent a considerable departure from the ideal conditions affects the net behavior, and second, how this situation may be partly improved by a suitable choice of terminal resistances.

4. The behavior in the transmission range. This discussion may best be given by means of a numerical example. Suppose we choose for simplicity the following component reactances

$$\left.\begin{aligned} z_1 &= jL\omega \\ z_2 &= \frac{1}{jC\omega} \end{aligned}\right\}. \tag{648}$$

Since the transmission range according to (641) lies in the vicinity of the zeros of z_1, we see that the reactances (648) will give rise to a low-pass class. This we know anyway from our experience with artificial and lump-loaded lines.

The condition (641) then becomes

$$\left.\begin{aligned} -1 \leq -\frac{LC\omega^2}{4} &\leq 0 \\ \text{or} \qquad 0 \leq \omega &\leq \frac{2}{\sqrt{LC}} \end{aligned}\right\}. \tag{649}$$

Let us choose

$$L = \frac{1}{\pi};\ C = \frac{10^{-6}}{\pi}. \tag{650}$$

so that we have for the cut-off frequency

$$\omega_c = \frac{2}{\sqrt{LC}} = 2\pi \cdot 1000. \tag{651}$$

In order to form the expression (637) for the characteristic impedance we note that

$$\sqrt{z_1 z_2} = \sqrt{\frac{L}{C}} = 1000,$$

and

$$\frac{z_1}{4z_2} = -\frac{LC\omega^2}{4} = -\left(\frac{\omega}{\omega_c}\right)^2.$$

Thus we have in this case

$$Z_T = 1000\sqrt{1 - \left(\frac{\omega}{\omega_c}\right)^2}. \tag{652}$$

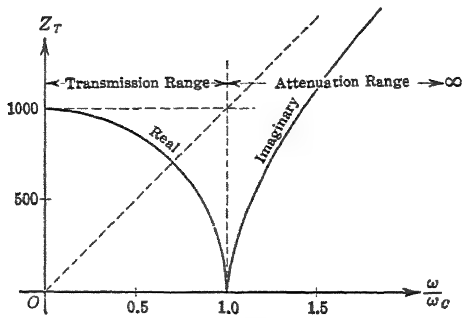


FIG. 103.—Mid-series characteristic impedance of low-pass structure according to (652).

A plot of this is shown in Fig. 103. We see that in the transmission range the characteristic impedance varies from the value

$$\sqrt{z_1 z_2} = 1000$$

to zero. It is real in this interval. In the attenuation range the impedance becomes a pure imaginary quantity and varies from zero to infinity.

It is thus not possible to have anything like good terminal conditions, even in the pass band. In order to see what a considerable degree of mismatch at the terminals will do, let us deliberately make things even worse by taking

$$Z_S = 1500; \; Z_R = 500. \tag{653}$$

Finally let us consider four T-sections in cascade so as to have a sufficient number of resonant frequencies to make the situation interesting.

For the determination of these frequencies we use (644), which becomes

$$\omega = \omega_c \sin \frac{\nu\pi}{16}. \tag{654}$$

Since the sine-function traverses its complete range of positive values in the interval 0 to $\pi/2$, we see that the range

$$\nu = 0, 1, 2, \cdots, 7, 8$$

will give all the desired frequencies. These are tabulated below with even and odd integer values in separate columns.

ν	ω	ν	ω
0	0	1	$2\pi\cdot 195$
2	$2\pi\cdot 382$	3	$2\pi\cdot 555$
4	$2\pi\cdot 707$	5	$2\pi\cdot 830$
6	$2\pi\cdot 923$	7	$2\pi\cdot 980$
8	$2\pi\cdot 1000$		

Substituting the values (653) into (643) we get for the locus of ν-even

$$\left|\frac{E_n}{E_g}\right|_{\nu\text{-even}} = 0.25. \tag{655}$$

Using (652) in (643a) we have for the locus of ν-odd

$$\left|\frac{E_n}{E_g}\right|_{\nu\text{-odd}} = \frac{0.5\sqrt{1-\left(\dfrac{\omega}{\omega_c}\right)^2}}{1.75-\left(\dfrac{\omega}{\omega_c}\right)^2}. \tag{656}$$

For the evaluation of the cut-off ratio (647) we have

$$|z_2|_c = \frac{1}{C\omega_c} = \frac{1}{2}\sqrt{\frac{L}{C}} = 500,$$

so that with $n = 4$

$$\left|\frac{E_n}{E_g}\right|_c = 0.25(1+9)^{-\frac{1}{2}} = 0.079. \tag{657}$$

The resulting behavior in the pass band is now found by plotting (655) and (656), and marking points on these corresponding to the frequencies given in the above table, with the exception of the cut-off frequency at which the ratio is separately given by (657). A zigzag line joining these points represents the desired behavior. This is shown in Fig. 104.

An interesting point is the distribution of resonant frequencies in the pass band. These become more closely spaced as the cut-off point is approached. The number of these frequencies increases with the number of cascaded sections, as may be expected. If we consider the limiting case of an infinite number of sections, the number of resonant frequencies in the pass band becomes infinite and the cut-off becomes a point of condensation for these frequencies. We then have a condition of continuous resonance, and since the terminal impedances are then virtually equal to the characteristic impedance, the loci of extremal values coincide in the constant value of $\frac{1}{2}$. This limit, therefore, represents the ideal condition.

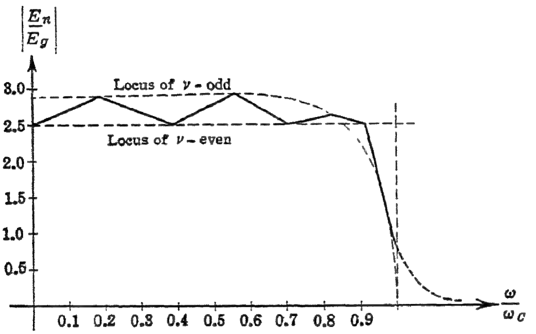


FIG. 104.—Response ratio showing behavior in the transmission range of the four-section low-pass filter with the parameter values (650) and terminal resistances (653).

The condition for which Fig. 104 is drawn was deliberately chosen with fairly poor terminal conditions. Yet the resulting behavior is not at all bad.

5. The behavior in the attenuation range. In the attenuation range it is more appropriate to discuss the logarithmic ratio of E_g to E_n. Since the phase function γ_2 is equal to $\pm\,\pi$ in this region, the relation (639) yields

$$\ln\left|\frac{E_g}{E_n}\right| = \ln\left(\frac{Z_S + Z_R}{Z_R}\right) + n\gamma_1 - \ln\left|1 - r_S r_R\right| + \ln\left|1 - r_S r_R e^{-2n\gamma_1}\right|. \quad (658)$$

The first term in this expression needs no comment. The second, which involves the attenuation function, may be calculated in a large number of ways owing to the fact that the inverse hyperbolic sine in (637) can be transformed into many other forms. Of these the loga-

rithmic form is perhaps best for numerical calculation. We shall, therefore, discuss this.

Applying the formula for the hyperbolic sine of the sum of two angles, and noting that $\gamma_2 = \pm\pi$, we have by (637)

$$\sqrt{\frac{z_1}{4z_2}} = \sinh\left(\frac{\gamma_1 \pm j\pi}{2}\right) = \pm j \cosh\frac{\gamma_1}{2}. \tag{659}$$

It is now convenient to make the substitution

$$\sqrt{\frac{z_1}{4z_2}} = \pm jx. \tag{660}$$

The usefulness of this will be appreciated throughout all the following discussions on the subject of conventional filter theory. For the time being the reader may content himself with the mere formality of this step. We thus have with the use of the elementary notions of hyperbolic functions

$$\left.\begin{aligned} \cosh\frac{\gamma_1}{2} = x;\ \sinh\frac{\gamma_1}{2} = \sqrt{x^2-1} \\ e^{\pm\frac{\gamma_1}{2}} = \left(x \pm \sqrt{x^2-1}\right) \end{aligned}\right\}, \tag{661}$$

so that

$$n\gamma_1 = \ln\left(x + \sqrt{x^2-1}\right)^{2n}. \tag{662}$$

This together with (660) is a convenient form for the calculation of the second and usually most important term in (658).

The next term, which involves the reflection coefficients and is designated as the **reflection loss,** requires some additional remarks. First of all we wish to point out that the characteristic impedance, given by (637), is a pure imaginary in any attenuation range when the structure is non-dissipative. According to (641) the ratio $z_1/4z_2$ is either less than minus one or greater than zero. In the former case, the product z_1z_2 is real, so that the first radical in (637) is real while the second is imaginary. On the other hand, if $z_1/4z_2$ is positive, then the product z_1z_2 is negative and the first radical in (637) is imaginary while the second is real. For the following discussion we may simplify matters if we let

$$\sqrt{z_1z_2} = R. \tag{663}$$

This expresses the fact that the component reactances are inverse networks with respect to R^2. In the majority of cases R is a real positive constant, as in the above numerical example. The following treatment applies particularly to this case. With (660) the characteristic impedance (637) may then be written

$$Z_T = jR\sqrt{x^2-1}. \tag{664}$$

Since the terminal impedances Z_S and Z_R are assumed to be real constants, the reflection coefficients (638) become quotients of conjugate complex numbers. If we let

$$\left.\begin{aligned} u_S &= \frac{Z_T}{jZ_S} = \frac{R}{Z_S}\sqrt{x^2-1} \\ u_R &= \frac{Z_T}{jZ_R} = \frac{R}{Z_R}\sqrt{x^2-1} \end{aligned}\right\}, \tag{665}$$

these become

$$r_S = \frac{1-ju_S}{1+ju_S}; \quad r_R = \frac{1-ju_R}{1+ju_R},$$

so that we have

$$|1 - r_S r_R| = \frac{2(u_S + u_R)}{\sqrt{(1+u_S^2)(1+u_R^2)}} = \rho, \tag{666}$$

where the symbol ρ is introduced as an abbreviation.

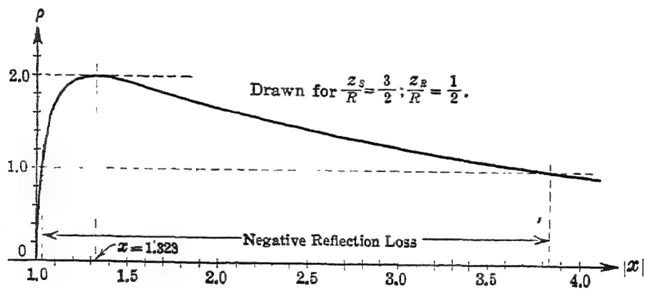


FIG. 105.—The reflection loss factor as a function of the normalized frequency $x = \omega/\omega_c$.

The factor ρ as a function of x is shown plotted in Fig. 105 for the range $|x| > 1$. Within this range the function passes from zero to a maximum value of 2 and back to zero. The maximum occurs where $u_S u_R = 1$, which corresponds to

$$|x| = \left(1 + \frac{Z_S Z_R}{R^2}\right)^{1/2}. \tag{667}$$

Thus there is a finite region where ρ lies between 1 and 2. Here the logarithm of ρ is positive, which makes the reflection loss negative; i.e., the effect of reflections in this portion of the attenuation range is to introduce a gain. The maximum value of this gain is, however, limited to $\ln 2 = 0.693$ napier $= 6.02$ decibels. For $|x| = \infty$ the reflection loss is infinite. At the point $|x| = 1$ the reflection loss is also infinite, but here $\gamma_1 = 0$ so that the last two terms in (658) are both infinite and

opposite. Their combined effect is finite, as already determined for the transmission range, and is evaluated by means of (647).

The last term in (658) is called the **interaction loss,** probably because it involves not only the terminal conditions at both ends but the propagation function as well. Whereas this term characterizes the behavior in the transmission range, it is of very little importance beyond the cut-off on account of the exponential factor $\epsilon^{-2n\gamma_1}$. With the help of (666) we find for the attenuation range

$$|1 - r_S r_R e^{-2n\gamma_1}| = \{1 - 2e^{-2n\gamma_1}(1 - \rho^2/2) + e^{-4n\ {}_1}\}^{1/2}. \tag{668}$$

Where the attenuation is at all appreciable the last term in this expression may be dropped and the second considered small compared to unity. Thus we get for the interaction loss

$$\ln|1 - r_S r_R e^{-2n\gamma_1}| \cong \ln\{1 - e^{-2n\gamma_1}(1 - \rho^2/2)\} \cong - e^{-2n\gamma_1}(1 - \rho^2/2)\cdot \tag{669}$$

The relations (661), (662), (666), and (669) evaluate the last three terms in (658) so that the variation of the net logarithmic voltage ratio in the attenuation range is readily calculated. The Fig. 105 for ρ is drawn for the numerical example stated above.

Let us illustrate the use of these results in the low-pass filter structure discussed in the previous section. Here we have

$$R = 1000;\ x = \frac{\omega}{\omega_c}\cdot \tag{670}$$

With (653) the relations (665) are

$$u_S = \frac{2}{3}\sqrt{x^2 - 1};\ u_R = 2\sqrt{x^2 - 1}. \tag{671}$$

Assuming a series of values for x such as $x = 1.1, 1.2, 1.3, \ldots$ the corresponding values for the most important term (662) are readily calculated. Likewise a series of u_S and u_R values may be determined and from these the terms (666) and (669) evaluated. For example, for $x = 1.1$ we have

$$\begin{aligned}
x \pm \sqrt{x^2 - 1} &= 1.558;\ 0.642 \\
e^{n\gamma_1} = (x + \sqrt{x^2 - 1})^8 &= 34.9 \\
4\gamma_1 = \ln 34.9 &= 3.55 \text{ napiers} \\
&= 30.9 \text{ decibels} \\
u_S = 0.305;\ u_R &= 0.916 \\
\rho = 1.72;\ \ln 1.72 &= 0.544 \text{ napier} \\
e^{-2n\gamma_1} = (x - \sqrt{x^2 - 1})^{16} &= 8.29\cdot 10^{-4} \\
-e^{-2n\gamma_1}\left(1 - \frac{\rho^2}{2}\right) &= 8.29\cdot 10^{-4}\cdot 0.48 = 3.98\cdot 10^{-4}.
\end{aligned}$$

Hence the last three terms in (658) in napiers are

$$3.556 - 0.544 + 3.98 \cdot 10^{-4},$$

or in decibels

$$30.9 - 4.72 + 3.45 \cdot 10^{-3}.$$

The last term is obviously negligible. It will be even more so for larger values of x. The second-last term, however, is not negligible in this case, although it is considerably smaller than the $n\gamma_1$ term.

The first term in (658) is a loss which is strictly not chargeable to the filter proper since it is present even though the terminal impedances are directly connected. The ratio

$$\frac{E_n'}{E_n} = \frac{Z_R}{Z_S + Z_R} \cdot \frac{E_g}{E_n}$$

is the **insertion ratio** of which we spoke in connection with the long-line problem. It is the ratio of the received voltage without and with the filter. The logarithm of this ratio is given by the last three terms in (658).

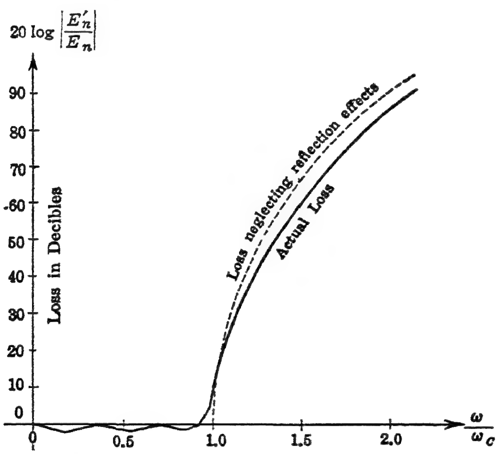


FIG. 106.—Insertion loss of the four-section low-pass filter.

Fig. 106 shows a plot of the insertion loss in decibels versus frequency for the numerical example just discussed. The points of this curve in the transmission range are taken from the plot of Fig. 104 calculated in the previous section. The dotted curve in Fig. 106 is for the attenuation term $n\gamma_1$ alone. The addition of this curve brings out the effect of improper terminations since the $n\gamma_1$ curve is that which would result for the net insertion loss if the filter were terminated in its characteristic impedance.

This figure shows that improper terminations have the effect of making the rise in the net attenuation less abrupt beyond the cut-off point as well as to introduce some attenuation within the pass band. Near the cut-off this attenuation is positive. At frequencies below the cut-off the reflections cause small negative losses. This corresponds to

the fact that in Fig. 104 the locus of ν-odd lies above that for ν-even over part of the pass band, so that at these frequencies the voltage ratio is higher with the filter than without it. This is merely a resonance phenomenon and not at all out of the ordinary. It is essential to note that this condition can occur only when $Z_S \neq Z_R$ as was pointed out earlier with regard to Fig. 102

An important point to note from Fig. 106 is the fact that, although the terminal resistances for this example were deliberately chosen so as to disagree considerably with Z_T, the reflection effects are not outstanding. It should be remembered, of course, that the reflection loss is independent of the number of cascaded sections and hence becomes relatively more noticeable as this number is decreased.

6. Adjustment for best terminal conditions. We shall now discuss some of the steps which may be taken with the view toward improving the terminal conditions within the transmission range. In the attenuation range nothing can be done toward bringing about a better agreement between the characteristic impedance and the terminal resistances primarily because the former is a pure imaginary. In the transmission range, however, the situation can be somewhat improved by means of the following procedure.

Let us consider the case $Z_S = Z_R$ for which the loci of extremal values become tangent at one point but do not cross. This was discussed in connection with Fig. 102. If we use the substitutions (660) and (663) in the expression for the characteristic impedance (637), the latter becomes

$$Z_T = R\sqrt{1 - x^2}. \tag{664a}$$

This is really the same as the form (664), but more applicable to the transmission region where x is less than unity.

We shall consider only the case where R is a real positive constant. Our object is to see whether, by a proper choice of the common terminal resistance relative to R, the loci (643) and (643a) can be made more nearly coincident over an appreciable part of the transmission region. In order to investigate this we may form the ratio of (643a) to (643) and then see how closely the resulting function may be made to approximate a constant value. Calling the common terminal resistance $R_t = Z_S = Z_R$, and using (664a), we have for this ratio

$$\left|\frac{E_n}{E_g}\right|_{\nu\text{-odd}} \Big/ \left|\frac{E_n}{E_g}\right|_{\nu\text{-even}} = f(x) = \frac{2RR_t\sqrt{1 - x^2}}{R_t^2 + R^2(1 - x^2)}. \tag{672}$$

The transmission range is in general that for which $-1 \leq x \leq 1$; but

since $f(x)$ is an even function of x it is necessary to study this function only for the range $0 \leq x \leq 1$.

Fig. 107 shows a plot of $f(x)$. The function has a minimum at $x = 0$, and a maximum at

$$x_{\max} = \sqrt{1 - \left(\frac{R_t}{R}\right)^2}. \tag{673}$$

The minimum value is

$$f_{\min} = f(0) = \frac{2RR_t}{R^2 + R_t^2}, \tag{674}$$

while the maximum value is unity. The *tolerance* we shall define as the maximum deviation from the average which is in this case

$$\epsilon = \frac{1 - f_{\min}}{1 + f_{\min}}$$

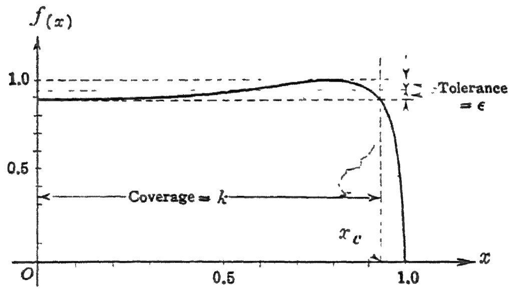


FIG. 107.—The ratio (672) as a function of the normalized frequency.

Substituting from (674) this gives

$$\epsilon = \left(\frac{1 - \dfrac{R_t}{R}}{1 + \dfrac{R_t}{R}}\right)^2. \tag{675}$$

This tolerance is not exceeded over a frequency range extending from $x = 0$ to that point where the function again has the same value as f_{min}. This range we shall call the *coverage* and denote by the letter k. It is equal to the positive root of the equation

$$f(x) = f(0)$$

which is

$$k = \sqrt{1 - \left(\frac{R_t}{R}\right)^4}. \tag{676}$$

It is interesting to note that the tolerance and coverage are both functions of the ratio of terminal resistance to R. Fig. 108 shows a plot of ϵ and k versus this ratio. We see that the coverage decreases with the tolerance so that the choice of the resistance ratio will have to be based on a compromise. However, the tolerance decreases very rapidly with the ratio R_t/R while the coverage remains nearly unity for a considerable range of values. Thus quite good combinations of ϵ and k may be had. Fig. 107, for example, is plotted for $R_t/R = 0.6$ which gives $\epsilon =$ 6.25 per cent, $k =$ 93.3 per cent. $R_t/R = 0.7$ will give $\epsilon =$ 3.1 per cent and $k =$ 87.2 per cent, and so forth.

Thus reasonably good behavior over the pass band may be expected if the terminal resistances are properly chosen relative to the quantity R. It should be pointed out in this connection, however, that this process in no way eliminates reflections at the terminals of the structure. The latter are still present even though a fairly favorable behavior has been obtained. In cases where the latter is the only criterion, the problem may be considered solved provided the resulting behavior is satisfactory. When filters are used in conjunction with transmission lines, however, reflections may still cause serious trouble in the way of echoes and interference. Then the situation must be met in a more satisfactory manner. How this may be done will be taken up in the following chapters.

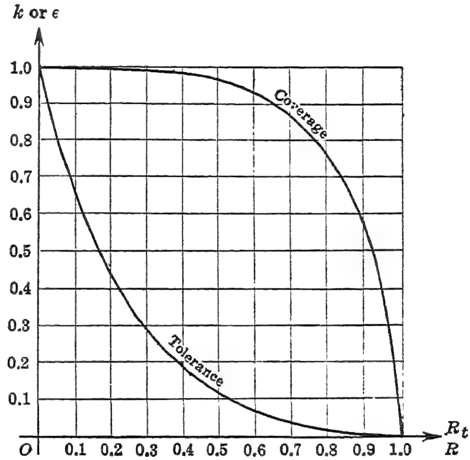


FIG. 108.—The relation between tolerance and coverage for the function of Fig. 107.

PROBLEMS TO CHAPTER VIII

8-1. A uniform ladder structure consists of a cascade of identical T-sections as illustrated in Fig. 101 for which:

$$z_1 = 1/jC\omega; \; z_2 = jL\omega.$$

Show that the network has a theoretical range of free transmission for all frequencies above $\omega_c = 1/2\sqrt{LC}$. If the structure consists of five sections in cascade, determine the frequencies at which the extremal values in the voltage response ratio occur, and determine the value of the common terminal resistance for which the maximum

average deviation between the maximum and minimum response does not exceed 5 per cent over the practical transmission range. What are the frequency limits within which this maximum deviation is not exceeded? Choose $\omega_c = 2\pi \cdot 1000$, and $\sqrt{L/C} = 1000$ ohms.

8–2. For the structure of Problem 8–1 with its terminal resistances, determine the logarithmic insertion loss ratio for the voltage throughout the attenuation range. Plot this and also the corresponding curve obtained by neglecting the reflection effects. Do the same for a single T-section, and note the relatively increased effect of the reflection and interaction losses.

8–3. How is the procedure in the determination of the results called for in Problems 8–1 and 8–2 modified if the structure is composed of symmetrical Π- instead of T-sections?

8–4. Repeat the essential features involved in Problems 8–1 and 8–2 for a chain of tuned coupled circuits which may be specified as a cascade of symmetrical T- or Π-sections for which

$$z_1 = j(L - M)\omega + \frac{1}{jC\omega}; \; z_2 = jM\omega.$$

Let the transmission range be defined by $\omega_1 < \omega < \omega_2$. Show that in general

$$\frac{4M}{L - M} = \frac{2w}{\omega_1}\left(1 + \frac{w}{2\omega_1}\right),$$

where $w = \omega_2 - \omega_1$, and $\omega_2 = 1/\sqrt{(L - M)C}$. Determine and plot the mid-series characteristic impedance. Consider in detail only the case for which $\frac{w}{\omega_1} \ll 1$, and show that then the characteristic impedance has a maximum value of $M\omega_2$ in the transmission range. For a single mid-series section show that the results agree with those found for a pair of tuned coupled circuits as discussed in detail on pages 327–335 of Vol. I.

8–5. On the basis of the results of Problem 8–5, design a selective network consisting of four cascaded tuned coupled circuits to work into and out of 1,000 ohms with a band width of 10^4 cycles per second located at 10^6 cycles per second.

8–6. Consider a cascade of four identical T-sections of the constant-k band-pass type whose series and shunt impedances are derived on pages 318 and 319 in the next chapter. Let the cut-off frequencies be 1000 and 2000 cycles per second, and let the nominal value R of the characteristic impedance be 1000 ohms. Plot the magnitude of the voltage ratio over the transmission range for values of identical terminal resistances equal to 700, 1000, and 2000 ohms.

8–7. For the structure of Problem 8–6 determine and plot the insertion loss over the ranges 0–1000 and 2000–4000 cycles per second for terminal resistance values of 1000 ohms. Plot also the attenuation loss alone, and note the effect of imperfect matching at the terminals.

8–8. A ladder network consists of a cascade of two symmetrical T-sections of the type considered in section 4 with the parameter values given by eq. (650). A voltage E_0, having a frequency of 707 cycles per second, is impressed directly to the sending end ($Z_S = 0$). Determine the magnitude and phase (with respect to E_0) of the midpoint voltage E_1 and the output voltage E_2 for the following load impedances:

$$Z_R = 0; \; Z_R = 707; \; Z_R = 1000; \; Z_R = j1000; \; Z_R = \infty.$$

CHAPTER IX

CONVENTIONAL FILTER THEORY

1. Basic design considerations. The conventional filter theory utilizes the properties of the uniform ladder structure as a point of departure, and, by a process of repeated adjustments of the same nature, arrives at a sequence of operations whereby the more obvious requirements of a selective system may be attained to any desired degree of approximation consistent with economic considerations. Since it makes the uniform ladder structure basic, the ultimate results obtainable by the method are limited to the ability of such a structure in meeting the requirements set. The question, therefore, is left unanswered as to whether another structure may not meet the same requirements with better economy or perhaps with less effort in the design process. Logically the ideal design process ought not to limit itself to any basic type of structure[1] but rather arrive at a most economical structure from the required characteristics sheerly by reason of its own completeness. Our present method of filter design begins with the analysis of a particular form of structure, and, out of a classification of its modes of behavior, those network combinations are selected which most nearly meet the desired characteristics.

Before getting into the mathematical details, let us reflect for a moment upon that which we wish to accomplish. The reader appreciates by this time that the behavior of a four-terminal network is quite effectively characterized by its propagation and characteristic impedance functions. This is, of course, by no means the only way in which to describe the network's characteristics, but it has shown itself to be very effective in this case, particularly when the behavior of the filter is subsequently to be considered in conjunction with long lines or structures of a similar nature.

Assuming, then, that the behavior is to be described in terms of the propagation and characteristic impedance functions, what in general are the requirements or ideal characteristics which these are to exhibit in a filter network? Since the network is supposed to isolate regions of transmission and attenuation sharply, the most obvious requirement is that the real part of the propagation function (attenuation function)

[1] Unless such a structure can be shown to be potentially equivalent to the most general form of four-terminal network. This point is treated in more detail in the next chapter.

shall be as small as possible (preferably zero) in a transmission region, and as large as possible in an attenuation region, the transition between the two being as abrupt as possible. In a practical case a minimum value within the attenuation region as well as a definite rate of transition may be prescribed. The attenuation function thus is the major factor in controlling the selective properties of the network.

In addition to being selective, the network must transmit with a minimum of distortion all frequencies within its pass band. This imposes further requirements with which we are familiar from our study of long-line behavior. In the transmission range the attenuation function must be constant and the phase function must be linear with respect to frequency, both within maximum tolerances set by the specific problem in hand. In addition to this the characteristic impedance should equal the terminal impedances to within a prescribed tolerance. This is necessary in order to suppress reflections which otherwise not only would introduce amplitude and phase distortion, but also might produce interference and echoes. Since the terminal impedances are usually resistances, the characteristic impedance is required to be a real constant within a transmission range.

From our study of the uniform ladder structure we can appreciate that several obstacles must be overcome before this type of network can be made to yield such results. One of these presents itself at the outset from a further study of the expressions for the propagation and characteristic impedance functions. We repeat them for convenience:

$$\left.\begin{aligned} \gamma &= 2\sinh^{-1}\sqrt{\frac{z_1}{4z_2}} \\ Z_{\mathrm{T}} &= \sqrt{z_1 z_2}\sqrt{1+\frac{z_1}{4z_2}} \end{aligned}\right\}. \tag{677}$$

It is apparent that the propagation function depends only upon the ratio of the component impedances, whereas the characteristic impedance involves both the product and the ratio. This means that these two functions cannot be adjusted in any simple manner to meet specifications *independently* of each other. Thus while we may be adjusting the component impedances to meet the requirements of γ, the characteristic impedance function will be affected. Hence a separate adjustment for each function is difficult. This fact does complicate the procedure somewhat as we shall see later. For the time being we wish to concentrate our attention more specifically upon the propagation function in order to get a clearer picture of its dependence upon the component impedances.

Let us not make the assumption initially that we are treating a non-

dissipative structure. Although we shall be obliged to do this later in developing design formulae, we wish to give a more general discussion first so that the exact performance of a physical structure may be determined if necessary. Consider again the change of variable

$$\sqrt{\frac{z_1}{4z_2}} = \pm jx, \tag{678}$$

but with the understanding that x may in general be complex, namely, that we can further write

$$x = \xi + j\eta. \tag{678a}$$

The first equation (677) then yields

$$\pm \xi \pm j\eta = \cosh\frac{\gamma_1}{2}\sin\frac{\gamma_2}{2} - j\sinh\frac{\gamma_1}{2}\cos\frac{\gamma_2}{2},$$

and equating reals and imaginaries we have

$$\left.\begin{aligned} \xi &= \pm\cosh\frac{\gamma_1}{2}\sin\frac{\gamma_2}{2} \\ \eta &= \mp\sinh\frac{\gamma_1}{2}\cos\frac{\gamma_2}{2} \end{aligned}\right\}. \tag{679}$$

For a given structure, ξ and η are known functions of frequency, and the problem is to determine γ_1 and γ_2 from (679). This is easily done as follows: Dividing the first equation by $\cosh\frac{\gamma_1}{2}$, the second by $\sinh\frac{\gamma_1}{2}$, squaring, and adding, we have

$$\frac{\xi^2}{\cosh^2\frac{\gamma_1}{2}} + \frac{\eta^2}{\sinh^2\frac{\gamma_1}{2}} = 1. \tag{680}$$

On the other hand, if we divide the first equation by $\sin\frac{\gamma_2}{2}$, the second by $\cos\frac{\gamma_2}{2}$, square both, and subtract the second from the first, we get

$$\frac{\xi^2}{\sin^2\frac{\gamma_2}{2}} - \frac{\eta^2}{\cos^2\frac{\gamma_2}{2}} = 1. \tag{681}$$

These equations in γ_1 and γ_2 respectively may most conveniently be solved graphically. Considering ξ and η as the independent variables, equation (680) is the equation of an *ellipse* in normal form with semi-major axis $= \cosh\frac{\gamma_1}{2}$, semi-minor axis $= \sinh\frac{\gamma_1}{2}$ and foci at $\xi =$

± 1. For various values of γ_1 a family of concentric confocal ellipses may be drawn in the complex plane $x = \xi + j\eta$. For any given x the corresponding γ_1 may then be read graphically.

The relation (681), on the other hand, represents the equation of a *hyperbola* in normal form with semi-transverse axis $= \sin\frac{\gamma_2}{2}$, semi-conjugate axis $= \cos\frac{\gamma_2}{2}$, and foci at $\xi = \pm 1$. For various values of γ_2 this relation, therefore, will yield a family of confocal hyperbolae,

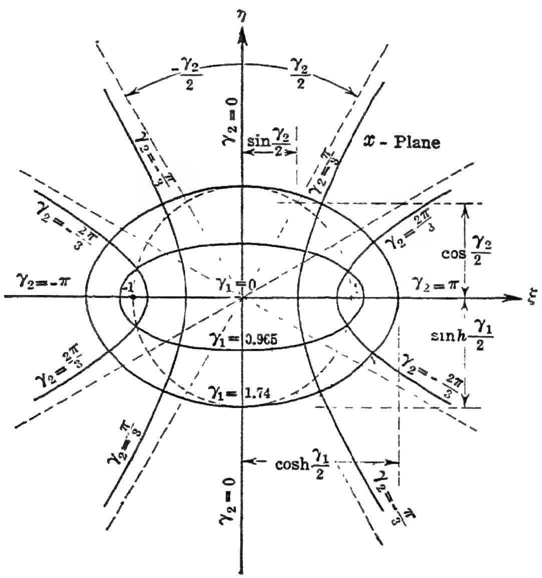


FIG. 109.—Orthogonal families of constant attenuation and constant phase loci for the uniform ladder network in terms of the frequency function x as defined by eq. (678).

the foci being coincident with those for the ellipses (680). The ellipses and hyperbolae may easily be shown to form *orthogonal* families. This is so because

$$\left(\frac{d\xi}{d\eta}\right)_{\gamma_1 = \text{const}} = -\left(\frac{d\eta}{d\xi}\right)_{\gamma_2 = \text{const}}$$

as may be verified readily from (679). The manner in which these families are constructed is illustrated in Fig. 109 which represents the complex x-plane. With a sufficient number of curves drawn, a chart of this kind may be used to determine γ_1 and γ_2 corresponding to any given complex x-value. In particular, the value $\gamma_1 = 0$ gives rise to the degenerate ellipse represented by the line joining the foci. In order to

be able to realize this condition, x must evidently be purely real; i.e., η must be zero. This in turn is possible only with pure reactive component impedances.

Similarly the values $\gamma_2 = \pm \pi$ give rise to the degenerate hyperbolae represented by the portions of the ξ-axis outside the interval -1 to $+1$. These also are realized only with non-dissipative structures. For these, x is either purely real or purely imaginary. The most common case is that for which x is purely real. This occurs when the component reactances are inverse with respect to a real positive constant. We wish to consider this case in some detail now.

2. The constant-k type. The so-called **constant-k type** of filter is one for which the component impedances are inverse functions of impedance product R^2 where R is a real positive constant. In order to indicate that the component impedances are those for this type it is customary to add the subscript k to the impedances and write the defining equation as

$$z_{1k}z_{2k} = R^2. \tag{682}$$

From this it follows that

$$\frac{z_{1k}}{4z_{2k}} = \frac{z_{1k}^2}{4R^2} = \frac{R^2}{4z_{2k}^2}, \tag{683}$$

so that the substitution (678) becomes

$$x_k = \frac{z_{1k}}{2jR}, \tag{684}$$

where the subscript k is again added to designate that this variable x is for the constant-k type.

For non-dissipative structures the variable x_k is thus seen to be a real function of frequency similar to the reactance function z_{1k}. When the latter is known, x_k is immediately given by (684). In this equation we need no longer retain the minus sign, for reasons which will become clear later

In the following discussion we shall treat the case of the non-dissipative structure only. The presence of resistances in practical structures is considered incidental; i.e., they have no part in the filter design theory. Their presence in actual structures usually modifies the behavior predicted on the non-dissipative basis only to a moderate degree, and this modification can always be determined by means of the method discussed in the previous section.[1]

[1] A more useful approximate method of considering the effect of incidental dissipation is discussed in section 13 of the following chapter where a fuller appreciation of how dissipation affects the design procedure in cases requiring more careful consideration of the problem of distortion will also be gained.

Since x_k is a real variable, the families of ellipses and hyperbolae of Fig. 109 are not very well suited for the determination of γ_1 and γ_2 corresponding to a given x_k. The complex plane is no longer necessary in this case, and we shall, therefore, discuss the real functions for γ_1 and γ_2 versus x_k. These are obtained from (679) by considering $\eta = 0$. Then we have *either*

$$\left.\begin{aligned}\gamma_1 &= 0\\ \text{whence}\qquad \gamma_2 &= 2\ \sin^{-1}x_k\end{aligned}\right\}, \tag{685}$$

or

$$\left.\begin{aligned}\gamma_2 &= \pm\,\pi\\ \text{whence}\qquad \gamma_1 &= 2\ \cosh^{-1}|x_k|\end{aligned}\right\}. \tag{686}$$

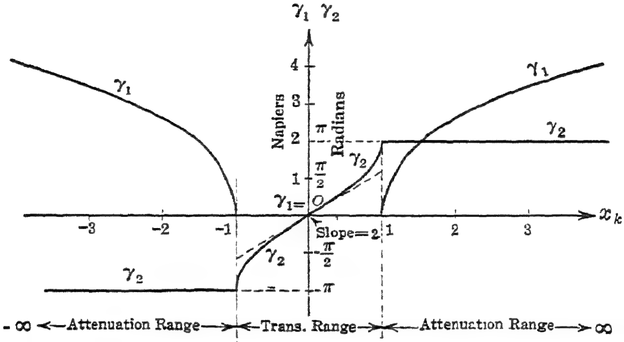


FIG. 110.—Dependence of the attenuation and phase functions upon the frequency function x_k for the non-dissipative constant-k filter as defined by eqs. (682) and (684).

The algebraic signs are justified on physical grounds in what is to follow.

The solutions (685) hold for the interval

$$-1 \leq x_k \leq 1, \tag{687}$$

and correspond to the transmission range. In Fig. 109 this corresponds to the ξ-axis between -1 and $+1$. The solutions (686), on the other hand, hold for

$$|x_k| \geq 1, \tag{687a}$$

and correspond to the attenuation range. Fig. 110 summarizes these solutions in both ranges. The negative branches of γ_1 are discarded because negative γ_1-values are impossible in a passive structure. The phase function γ_2 is restricted by the condition that its slope may never be negative. We shall see later in the discussion of the transient behavior of filters that the slope of the phase function equals the delay

time. Since the latter must always be positive, the restriction just stated follows.[1] In addition to this it should be borne in mind, of course, that γ_2 is determined only to within an additive integer number of 2π radians.

We turn now to a discussion of the characteristic impedance of the constant-k type. Here we wish to adopt the notation and terminology suggested by O. J. Zobel, who is principally responsible for the following development. In his earlier work[2] Zobel used the letter K for characteristic impedances, probably because the letter Z was already serving for a number of things. Later[3] he adopted W for this purpose as being more suitable, and we shall concur with him in this selection. Furthermore, it is necessary to distinguish between the characteristic impedances of T- and Π-sections, i.e., mid-series and mid-shunt sections, respectively. This distinction is made by adding the subscript *1* for mid-series and the subscript *2* for mid-shunt, which is in line with the subscript notation used for the series and shunt component reactances. Finally the subscript k is added to designate the constant-k type to which the characteristic impedances refer. Thus we write W_{1k} for the *mid-series* characteristic impedance, and W_{2k} for the *mid-shunt.*

Using (677) and recalling (427a), page 179, for the Π-section, we have

$$\left.\begin{aligned} W_{1k} &= R\sqrt{1 - x_k^2} \\ W_{2k} &= \frac{R}{\sqrt{1 - x_k^2}} \end{aligned}\right\}. \tag{688}$$

Plots of these are given in Fig. 111. Within the transmission region the functions are real; in the attenuation ranges they are purely imaginary. Where these functions are purely imaginary they must behave like physically realizable reactances and hence, according to what was said in Chapter V, they must have positive slope. This determines their algebraic signs which, on account of the radicals, may be either positive or negative.

From (688) it is apparent that

$$W_{1k} \cdot W_{2k} = R^2, \tag{689}$$

i.e., these are inverse functions with respect to R^2. R is called the

[1] This assumes that lead angles are defined as being negative and lag angles positive. Moreover, this restriction as to the positiveness of the slope of the phase characteristic applies only to non-dissipative networks, as will be discussed more fully in Chapter XI.

[2] O. J. Zobel, "Theory and Design of Uniform and Composite Electric Wave-Filters," B.S.T.J., January, 1923.

[3] O. J. Zobel, "Extensions to the Theory and Design of Electric Wave-Filters," B.S.T.J., April, 1931.

nominal value of the characteristic impedances. It is also interesting to note that if the ordinate R is chosen the same size as one unit on the x_k-scale, then W_{1k} is simply a semicircle within the interval $-1 \leq x_k \leq 1$. Thus neither W_{1k} nor W_{2k} is particularly constant throughout the transmission range, and hence reflection effects will be present to a certain extent, particularly in the vicinity of the cut-off points. The results of the preceding chapter have shown, however, that, to a first approximation, the behavior is still given by the propagation function alone. This fact enables us to proceed with the formulation of a design process which, for the moment, neglects reflection effects entirely.

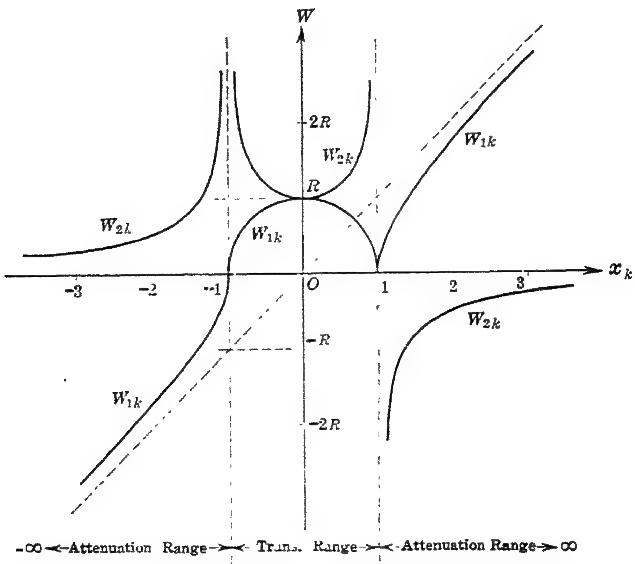


FIG. 111.—Dependence of the mid-series and mid-shunt characteristic impedances upon the frequency function x_k for the non-dissipative constant-k filter.

Later this question will be considered in more detail, and means will be developed to more adequately meet the design requirements.

3. Design procedure for the constant-*k* type. On the basis of the above preliminary remarks we are now in a position to formulate a definite procedure for the design of a filter of this type to meet any specifications regarding the allocation of transmission and attenuation regions in the frequency spectrum. The usual practice is to choose the nominal value of the characteristic impedance, i.e., R, equal to the resistance that the filter is to work into and out of. In the preceding chapter we saw that, with a mid-series termination, somewhat better

results are obtained when R is larger. Similarly, for a mid-shunt termination R ought to be chosen somewhat smaller, as may be appreciated from Fig. 111. Whatever is decided upon here will determine the value of R for any practical case. Hence we need concern ourselves, for the design, only with z_{1k}. When this is determined, z_{2k} may be found by reciprocation as discussed in section 8 of Chapter V.

The reactance z_{1k} is determined by means of the relation (684) together with the fact that a transmission range occurs where x_k lies between -1 and $+1$, as illustrated by Fig. 110. The points $x_k = \pm 1$, which correspond to

$$z_{1k} = \pm 2jR, \tag{690}$$

mark the *cut-off frequencies.* The point $x_k = 0$ corresponds to $z_{1k} = 0$, i.e., to a zero for the reactance z_{1k}. The latter, therefore, must have as many zeros as the filter is to have transmission regions. A zero of z_{1k} will not lie in the center of this region because z_{1k} is not a linear function of frequency in this vicinity, but it will be somewhere near the center at all events. Since a reactance function has either a zero or a pole at the origin and at infinity, that is, since z_{1k} cannot have a finite non-zero value at either of the extremities of the frequency spectrum, it follows that a pass band starting at the origin or terminating at infinity requires that z_{1k} have a zero at the origin or at infinity, respectively. These bands or regions we shall call **external** ones to distinguish them from **internal regions** for which both boundaries lie at finite non-zero frequencies.

Thus z_{1k} must have one zero for each internal pass band, and a zero at the origin or at infinity for an external pass band according to whether the latter is located at the origin or at infinity, respectively. This determines the structure of z_{1k}, the details of this process having been discussed in Chapter V. It might be well to recall for the reader, however, that any reactance function is determined by specifying the locations of its internal poles and zeros plus one additional piece of information which may be the value of the reactance at one other frequency. That is, the number of determining factors is equal to the number of internal zeros, plus the number of internal poles, plus one. This may also be stated as *twice the number of internal zeros, plus the number of external zeros (at 0 or ∞).* This number equals the least number of elements by means of which the reactance may be realized. From the above discussion, however, we see that this number coincides with the number of cut-off frequencies or boundaries between transmission and attenuation ranges. Hence, since equation (690) may be written for each boundary, we see that the number of independent equations thus

obtained just suffices for the determination of the elements in the structure of z_{1k}.

The procedure in any practical case is to set down the structure of z_{1k} first, say in the form of resonant components in parallel, one for each pass band, and label the inductances and capacitances, for example, as $L_1, L_3, \ldots L_{2\nu-1}$, and $C_3, C_5, \ldots C_{2\nu+1}$. This is illustrated in Fig. 112. The elements L_1 and $C_{2\nu+1}$ represent pass bands at the origin and at infinity, respectively. Here 2ν equals the *total* number of coils and condensers and hence also *the number of cut-off frequencies.* The next step is to write down z_{1k} as a function of ω and in terms of the L's and C's. Then write an equation of the form (690) at each of the specified cut-off frequencies, which we may label $\omega_{c1}, \omega_{c2}, \ldots \omega_{c2\nu}$. Since z_{1k} has positive slope, the minus sign in (690) is used at each left-hand boundary of a transmission region, and the plus sign at each right-hand boundary. This system of 2ν equations must then be solved simul-

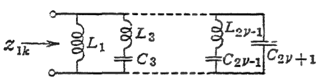


FIG. 112.—General form of the series reactance for the constant-k filter.

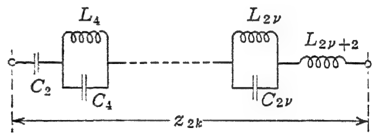


FIG. 112a.—General form of the shunt reactance for the constant-k filter.

taneously for the 2ν, L's, and C's. This is in general not a simple matter because the equations are not linear. Foster's reactance theorem cannot simplify matters because the locations of the poles and zeros are only approximately known. However, the problem is uniquely determined from the analytic side.

With z_{1k} determined, z_{2k} is its reciprocal with respect to R^2. For the network of Fig. 112 this gives the structure shown in Fig. 112a, for which we have

$$\frac{L_1}{C_2} = \frac{L_3}{C_4} = \cdots = \frac{L_{2\nu-1}}{C_{2\nu}} = \frac{L_4}{C_3} = \frac{L_6}{C_5} = \cdots = \frac{L_{2\nu+2}}{C_{2\nu+1}} = R^2. \quad (691)$$

Hence, with R given, z_{2k} is fully determined.

The structures thus obtained, of course, may be transformed into any equivalent ones according to the methods discussed earlier.

It is interesting to point out here that since z_{1k} and z_{2k} each contains as many elements as there are cut-off points, and furthermore, since a mid-series or mid-shunt section involves three component impedances, it follows that a single section of a constant-k filter requires a total of $3k$ elements where k equals the number of cut-off points. Each addi-

tional section, however, requires only two additional impedances, so that n sections in cascade require a total of $(2n + 1)k$ elements.

4. The use of half-sections. The discussion so far relates primarily to symmetrical T- or Π-sections, or a cascade of such sections. As we know from previous discussions, the cascading of sections multiplies the overall propagation function, but leaves the terminal conditions with regard to reflections unaltered. Furthermore, following the suggestion that the nominal impedance value R be made equal to or some factor times the resistance the filter is to work into and out of, it appears that the filter cannot be designed to work into and out of different terminal resistances. When this condition is desired in practice it is usually met by using a transformer at one end with suitable ratio. This procedure not only involves an added expense, but may lead to poor overall behavior unless the transformer is properly designed to have good transmission properties over the pass band of the filter. If reflections must be avoided for reasons other than those involved in the behavior of the filter itself, the use of a transformer is the simplest way of meeting the situation.[1] This also involves methods of improving the characteristic impedance behavior over the pass band, to which we shall come in a later section.

When only the behavior of the filter needs to be considered, a way of treating different terminal resistances in some cases is as follows. We saw in section 7 of the preceding chapter that a filter of the constant-k type behaves better when, at a mid-series end, the terminal resistance is made smaller than the nominal value R. This is summarized by Fig. 108 which shows the tolerance and coverage which may be expected of the loci of extremal values for various ratios of terminal to nominal resistances. If this problem had been treated for mid-shunt terminations, we would have found the same functional dependence for the tolerance and coverage but in terms of the inverse ratio of terminal to nominal resistances. This is to be expected since the inverse ratio at a mid-shunt end gives rise to the same magnitude for the reflection coefficient. The reader may readily verify this for himself, bearing in mind that the characteristic impedances at mid-series and mid-shunt ends are inverse functions with respect to R^2.

Hence if we use a structure with a mid-series termination at one end and a mid-shunt at the other, we may expect to get good results for a ratio of, say,

$$\frac{R_{t1}}{R} = 0.707$$

[1] Other methods of solving this problem by means of dissymmetrical structures will be discussed in section 11 of the following chapter.

at the mid-series end, and a ratio of

$$\frac{R}{R_{t2}} = 0.707$$

at the mid-shunt end, so that we then have

$$\frac{R_{t2}}{R_{t1}} = 2$$

for the ratio of terminal resistances. The simplest structure which offers such possibilities is the so-called **half-section** shown in Fig. 113. A cascade of an odd number of these on an image basis will present the same terminal possibilities but with proportionately increased attenuation and phase functions. This scheme is quite feasible, and we shall discuss it in more detail for several special cases.

The half-section is a dissymmetrical section. Our problem involves the cascade of an *odd* number of half-sections on the image basis. This

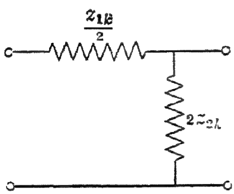


FIG. 113.—Half-section of the constant-k filter.

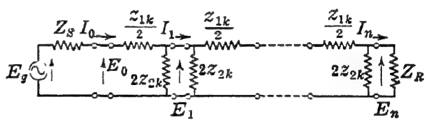


FIG. 114.—Cascade of an odd number of constant-k half-sections on the image basis.

was treated in the latter part of section 8 of Chapter IV, the results being given by equations (402) and (403), page 172. Let us consider n half-sections cascaded on this basis, working out of a resistance Z_S and into a resistance Z_R, with a mid-series termination at the left and a mid-shunt at the right. The ratio of interest shall be that for the voltage across Z_R to the generator voltage. This is illustrated in Fig. 114. Using (402) we then have

$$\frac{E_n}{E_g} = \sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}} \cdot \frac{Z_{I1}(1 + r_{R2})}{(Z_S + Z_{I1})(e^{n\theta} - r_{S1}r_{R2}e^{-n\theta})}. \tag{692}$$

The reflection coefficients are given by (399), the image impedances by (392) and (398), and $\mathfrak{D}$ and $\mathfrak{A}$ are two of the general circuit parameters (295), page 138. Noting that

$$\frac{1 + r_{R2}}{(Z_S + Z_{I1})} = \frac{Z_R(1 - r_{S1}r_{R2})}{Z_SZ_{I2} + Z_RZ_{I1}},$$

and by (398) that

$$\sqrt{\frac{\mathfrak{D}}{\mathfrak{A}}} = \sqrt{\frac{Z_{I2}}{Z_{I1}}},$$

we have for the magnitude of this ratio in the transmission range, i.e., for $\theta = j\theta_2$,

$$\left|\frac{E_n}{E_g}\right| = \frac{Z_R\sqrt{Z_{I1}Z_{I2}}}{Z_S Z_{I2} + Z_R Z_{I1}} \cdot \frac{|1 - r_{S1}r_{R2}|}{|1 - r_{S1}r_{R2}e^{-2jn\theta_2}|}. \tag{693}$$

Here θ is the image propagation function for the half-section and hence equal to $\gamma/2$ (γ being the propagation function for the symmetrical T- or Π-section). By (685) we have for the transmission range

$$\theta_2 = \sin^{-1} x_k. \tag{694}$$

Throughout this range the ratio (693) passes through maximum and minimum values, as discussed for a similar case in the preceding chapter. The extremal values occur very nearly for

$$e^{-2jn\theta_2} = \pm 1; \ \theta_2 = \frac{\nu\pi}{2n}; \ \nu = \begin{Bmatrix} \text{even} \\ \text{odd} \end{Bmatrix}. \tag{695}$$

The loci of extremal values are, therefore,

$$\left|\frac{E_n}{E_g}\right|_{\nu\text{-even}} = \frac{Z_R\sqrt{Z_{I1}Z_{I2}}}{Z_S Z_{I2} + Z_R Z_{I1}}, \tag{696}$$

and

$$\left|\frac{E_n}{E_g}\right|_{\nu-\text{odd}} = \frac{Z_R\sqrt{Z_{I1}Z_{I2}}}{Z_S Z_R + Z_{I1}Z_{I2}}. \tag{697}$$

In our case

$$Z_{I1} = W_{1k}; \ Z_{I2} = W_{2k}.$$

Using (688) we thus get

$$\left|\frac{E_n}{E_g}\right|_{\nu-\text{even}} = \frac{Z_R\sqrt{1 - x_k^2}}{Z_S + Z_R(1 - x_k^2)}, \tag{696a}$$

and

$$\left|\frac{E_n}{E_g}\right|_{\nu-\text{odd}} = \frac{RZ_R}{Z_S Z_R + R^2}. \tag{697a}$$

These are the loci of extremal values in the transmission range; i.e., for $-1 \leq x_k \leq 1$. The first of these loci varies with x_k; the second is constant. The problem is to determine Z_S, Z_R, and R such that (696a) approximates (697a) with a reasonable tolerance over a considerable portion of the band, and at the same time have $Z_S \neq Z_R$.

This is most conveniently done by forming the ratio of the loci and examining the ability of the resulting function to approximate unity in the prescribed range. Let

$$f(x_k) = \left|\frac{E_n}{E_g}\right|_{\nu-\text{even}} \Big/ \left|\frac{E_n}{E_g}\right|_{\nu-\text{odd}}$$

Then

$$f(x_k) = \frac{(Z_S Z_R + R^2)\sqrt{1 - x_k^2}}{R\{Z_S + Z_R(1 - x_k^2)\}}. \tag{698}$$

This function has a minimum value at $x_k = 0$, and a maximum value at

$$x_{k\max} = \pm\sqrt{1 - \frac{Z_S}{Z_R}}.$$

These values are

$$f_{\min} = \frac{Z_S Z_R + R^2}{R(Z_S + Z_R)}, \tag{699}$$

and

$$f_{\max} = \frac{Z_S Z_R + R^2}{2R\sqrt{Z_S Z_R}}. \tag{699a}$$

Here we may demand that the arithmetic mean value of the function be unity. This is expressed by

$$\tfrac{1}{2}(f_{\max} + f_{\min}) = 1, \tag{700}$$

and determines R in terms of Z_S and Z_R. If we write

$$\frac{Z_S}{Z_R} = r^2, \tag{701}$$

then (700) gives

$$R = \frac{2\sqrt{Z_S Z_R}(1 + r^2)}{(1 + r)^2}\left\{1 \pm \sqrt{1 - \frac{(1 + r)^4}{4(1 + r^2)^2}}\right\}. \tag{702}$$

Since both signs for the radical give positive results, either one may be used. This may be quite useful for the balance of the design. Note that the geometric mean of the roots (702) equals $\sqrt{Z_S Z_R}$.

The tolerance for $f(x_k)$ is expressed by

$$\epsilon = \frac{f_{\max} - f_{\min}}{f_{\max} + f_{\min}}. \tag{703}$$

Substituting (699) and (699a) into this we find

$$\epsilon = \left(\frac{1 - \sqrt{\dfrac{Z_S}{Z_R}}}{1 + \sqrt{\dfrac{Z_S}{Z_R}}}\right)^2. \tag{703a}$$

The coverage (symbol k) is that value of x_k for which $f(x_k)$ again equals $f_{\min}$; i.e., it is the root of the equation

$$f(x_k) = f_{\min}. \tag{704}$$

Substituting from above this gives

$$k = \sqrt{1 - \left(\frac{Z_S}{Z_R}\right)^2}. \tag{704a}$$

As functions of $\sqrt{Z_S/Z_R}$, (703*a*) and (704*a*) are identical in form to (675) and (676) of the preceding chapter. In Fig. 115 they are drawn as functions of Z_S/Z_R, which is more useful in this case. For example, for $Z_S/Z_R = \frac{1}{3}$, we find a 7 per cent tolerance over 94 per cent of the pass band. The corresponding values for R are found from (701) and (702). They are

$$R = 0\ 687 \cdot \sqrt{Z_S Z_R};\ 1.456 \cdot \sqrt{Z_S Z_R},$$

or

$$R = (1.190\, Z_S, 0.3967\, Z_R);\ (2.522\, Z_S, 0.841\, Z_R).$$

It is apparent that the half-section offers possibilities in the desired direction. Nevertheless it should be recognized that the chosen ratio materially affects the quality of the resulting behavior. Thus a ratio of $Z_R/Z_S = 2$ gives rise to a tolerance of about 3 per cent over 87 per cent of the pass band, while a ratio of $Z_R/Z_S = 4$ gives rise to a tolerance of 11 per cent over 97 per cent of the band. Ratios of Z_R/Z_S less than unity are obtained by reversing the half-section so that the input end becomes mid-shunt and the output mid-series. Fig. 115 may be used for this case by simply interchanging Z_S and Z_R, while equation (702) for R remains unchanged since it is unaffected by inverting r.

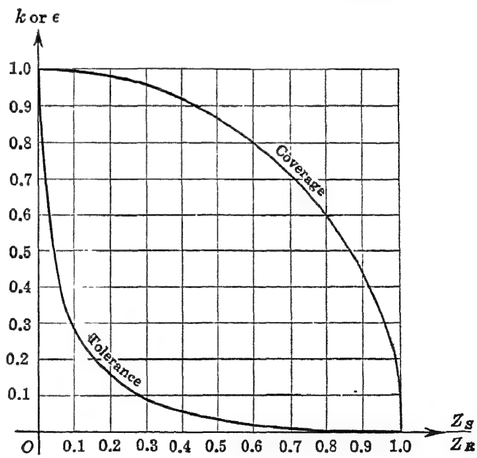


FIG. 115.—Dependence of tolerance and coverage of the function (698) upon the ratio of sending-to-receiving end resistances for the structure of Fig. 114.

It may occur in practice that a constant-k half-section is given, and it is desired to find the resistances Z_S and Z_R for which the behavior will

be most favorable. In this case R is given. Although Fig. 115 may be used to determine the ratio of Z_S to Z_R, this still leaves their values undetermined. However, the relation (700) must still be satisfied. After substituting from (699) and (699a) we find that this expression is symmetrical in the quantities R and $\sqrt{Z_S Z_R}$. Hence the solution for $\sqrt{Z_S Z_R}$ in terms of R and r is identical to (702). Namely, we have

$$\sqrt{Z_S Z_R} = \frac{2R(1+r^2)}{(1+r)^2}\left\{1 \pm \sqrt{1 - \frac{(1+r)^4}{4(1+r^2)^2}}\right\}. \tag{702a}$$

Thus, for a given R, $\sqrt{Z_S Z_R}$ may be determined after a ratio Z_S/Z_R has been chosen from Fig. 115 and the corresponding r found from (701). With the product and quotient of Z_S and Z_R known, these resistances are easily calculated. Since both signs in (702a) lead to positive values for $\sqrt{Z_S Z_R}$, either one may be used, so that we have a choice of values for which the resulting behavior will be equally good.

It should also be pointed out here that the ratio (693) becomes indeterminate at the cut-off points $x_k = \pm 1$, so that the locus (697a) for ν-odd does not hold at these boundaries. For the evaluation of the ratio at these points we express (693) in the trigonometric form and substitute

$$Z_{I1} = R \cos \theta_2; \; Z_{I2} = \frac{R}{\cos \theta_2}, \tag{705}$$

which are obtained with the use of (688) and (694), so that we have

$$\left|\frac{E_n}{E_g}\right| = \left|\frac{RZ_R}{\dfrac{RZ_S \cos n\theta_2}{\cos \theta_2} + RZ_R \cos n\theta_2 \cos \theta_2 + j(Z_S Z_R + R^2) \sin n\theta_2}\right|. \tag{706}$$

The value at the cut-off points is the limit for $\theta_2 \rightarrow \pm \frac{\pi}{2}$. Since n is *odd* this gives

$$\left|\frac{E_n}{E_g}\right|_c = \frac{RZ_R}{Z_S Z_R + R^2}\left\{1 + \left(\frac{nRZ_S}{Z_S Z_R + R^2}\right)^2\right\}^{-\frac{1}{2}} \tag{706a}$$

This takes the place of the value (697a) at the boundaries.

For certain problems we may be interested in the effect of an odd number of half-sections upon the *insertion* ratios. As has been mentioned previously, these are the ratios of output with and without the filter interposed between Z_S and Z_R. Without the filter we have

$$\left.\begin{aligned} E_n' &= \frac{E_g Z_R}{Z_S + Z_R} \\ I_n' &= \frac{E_g}{Z_S + Z_R} \end{aligned}\right\}. \tag{707}$$

Since $I_n Z_R = E_n$, we have

$$\frac{E_n}{E_n'} = \frac{I_n}{I_n'} = \frac{Z_S + Z_R}{Z_R} \cdot \frac{E_n}{E_g}, \tag{708}$$

so that the above results apply directly to the insertion ratios.

Another ratio which may be of interest is the volt-ampere ratio output-to-input. According to the notation given in Fig. 114 this is $E_n I_n / E_0 I_0$. Using (402) and (403) we find

$$\left|\frac{E_n I_n}{E_0 I_0}\right| = \frac{1 - r_{R2}^2}{|1 - r_{R2}^2 e^{-4jn\theta_2}|}. \tag{709}$$

Since this is an output-to-input ratio, it is independent of Z_S. Hence the ratio of terminal resistances does not affect it. Its behavior does depend, however, upon the resistance Z_R.

The extremal values occur for

$$e^{-4jn\theta_2} = \pm 1; \ \theta_2 = \frac{\nu\pi}{4n}; \ \nu = \begin{Bmatrix}\text{even}\\ \text{odd}\end{Bmatrix}.$$

The loci are

$$\left|\frac{E_n I_n}{E_0 I_0}\right|_{\nu\text{-even}} = 1, \tag{710}$$

and

$$\left|\frac{E_n I_n}{E_0 I_0}\right|_{\nu\text{-odd}} = \frac{2Z_R Z_{I2}}{Z_R^2 + Z_{I2}^2}. \tag{711}$$

For

$$Z_{I2} = \frac{R}{\sqrt{1 - x_k^2}},$$

we have for the second of these

$$\left|\frac{E_n I_n}{E_0 I_0}\right|_{\nu\text{-odd}} = \frac{2Z_R R\sqrt{1 - x_k^2}}{R^2 + Z_R^2(1 - x_k^2)}. \tag{711a}$$

This is the same as the function (672) except that the ratio R/Z_R replaces R_t/R. Hence the discussion given there applies here also, so that Fig. 108 may be used to find the tolerance and coverage in this case. At the cut-off frequencies, which correspond to even values of ν, (709) becomes indeterminate so that (710) does not apply. Evaluating in a manner similar to that used previously, we find for these boundary values

$$\left|\frac{E_n I_n}{E_0 I_0}\right|_c = \left(1 + \frac{n^2 R^2}{Z_R^2}\right)^{-\frac{1}{2}} \tag{712}$$

Before leaving this subject we wish to point out that what we have been denoting as the coverage is not to be interpreted as a percentage

of the width of the pass band in terms of frequency but merely as a percentage of the interval $-1 \leq x_k \leq 1$. The relation of this to the corresponding portion of the frequency band depends upon the function $x_k(\omega)$, which, according to (684), depends upon the allocation of transmission and attenuation regions in the frequency spectrum. This we wish to discuss in some detail for a few of the more common cases.

5. **Several common classes of constant-*k* filters.** In this section we wish to point out several details with regard to the four most common classes of filters of the constant-k type. These are the so-called **low-pass, high-pass, band-pass,** and **band-elimination** or **suppression** filters. The latter is sometimes also referred to as a *low-and-high-pass* filter. For convenience we shall designate these classes by **L.P., H.P., B.P.,** and **B.E.,** respectively. Although we have already discussed the **L.P.** filter to some extent, we include it here so that a comparison of the four classes may be made more easily.

For the **L.P.** *case we have*

$$z_{1k} = jL_1\omega; \; z_{2k} = \frac{1}{jC_2\omega}, \tag{713}$$

and since these are inverse functions with respect to R^2 it follows that

$$z_{1k}z_{2k} = \frac{L_1}{C_2} = R^2. \tag{714}$$

According to (684) we get

$$x_k = \frac{L_1\omega}{2R}.$$

At the cut-off frequency, which we denote by ω_c, this must equal $+1$, i.e.,

$$\frac{L_1\omega_c}{2R} = 1, \tag{715}$$

so that by substituting back in the previous equation we have

$$x_k = \frac{\omega}{\omega_c}. \tag{716}$$

Applying (685) and (686) we have

$$\left.\begin{aligned} \gamma_1 &= 0 \\ \gamma_2 &= 2\sin^{-1}\left(\frac{\omega}{\omega_c}\right) \end{aligned}\right\} \; (0 \leq \omega \leq \omega_c), \tag{717}$$

and

$$\left.\begin{aligned} \gamma_1 &= 2\cosh^{-1}\left(\frac{\omega}{\omega_c}\right) \\ \gamma_2 &= \pi \end{aligned}\right\} \; (\omega_c \leq \omega < \infty). \tag{718}$$

From (714) and (715) we have for the parameter values

$$L_1 = \frac{2R}{\omega_c}; \; C_2 = \frac{2}{R\omega_c}. \tag{719}$$

We see from these that a high nominal resistance R leads to a high L_1 and small C_2. In extreme cases the inductance may have more distributed capacitance than the value of C_2, which would render the design impractical. In such a case transformers must be used to reduce R. The cut-off frequency affects L_1 and C_2 in the same way. We have from (719)

$$\omega_c = \frac{2}{\sqrt{L_1 C_2}}. \tag{720}$$

For the **H.P.** *filter we have*

$$z_{1k} = \frac{1}{jC_1\omega}; \; z_{2k} = jL_2\omega, \tag{721}$$

so that

$$z_{1k}z_{2k} = \frac{L_2}{C_1} = R^2, \tag{722}$$

and

$$x_k = \frac{-1}{2RC_1\omega}.$$

The cut-off frequency for this case is a left-hand boundary for the transmission region. Hence

$$\frac{-1}{2RC_1\omega_c} = -1, \tag{723}$$

and thus

$$x_k = -\frac{\omega_c}{\omega}. \tag{724}$$

Note that for identical cut-off frequencies

$$x_{k\mathrm{H.P.}} = \frac{-1}{x_{k\mathrm{L.P.}}}. \tag{725}$$

With (724), γ_1 and γ_2 may be calculated easily from (685) and (686). From (722) and (723) we have for the parameter values

$$C_1 = \frac{1}{2R\omega_c}; \; L_2 = \frac{R}{2\omega_c}, \tag{726}$$

and hence

$$\omega_c = \frac{1}{2\sqrt{L_2 C_1}}. \tag{727}$$

The attenuation and phase functions for the various classes of filters are very conveniently found graphically. For this purpose the plots of Fig. 110 serve for all classes and may, therefore, be drawn up carefully over a considerable range. x_k is the parameter which varies from class to class according to the relation (684). Similarly, the curves of Fig. 111 for W_{1k} and W_{2k} may be drawn up carefully and made to serve for all classes. The reader should note that the curves of Figs. 110 and 111 show arithmetic symmetry with respect to the origin for x_k. They will not show this symmetry with respect to frequency, of course, since x_k is in general not a linear function of frequency. The **L.P.** filter is the only exception. In order to bring out the dependence of x_k upon frequency we have plotted this function for the **L.P.** and **H.P.** cases in Fig. 116. The curves are similar to those for the corresponding reactances z_{1k} as they should be since x_k is proportional to this reactance.

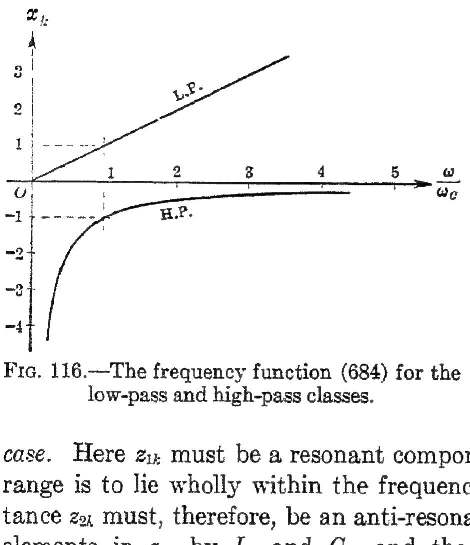


FIG. 116.—The frequency function (684) for the low-pass and high-pass classes.

Consider next the **B.P.** *case.* Here z_{1k} must be a resonant component because the transmission range is to lie wholly within the frequency spectrum. The shunt reactance z_{2k} must, therefore, be an anti-resonant component. Denoting the elements in z_{1k} by L_1 and C_1, and those in z_{2k} by L_2 and C_2, we have

$$\left.\begin{aligned} z_{1k} &= \frac{1 - L_1C_1\omega^2}{jC_1\omega} = \frac{jL_1}{\omega}\cdot(\omega^2 - \omega_0^2) \\ z_{2k} &= \frac{jL_2\omega}{1 - L_2C_2\omega^2} = \frac{\omega}{jC_2}\cdot\frac{1}{\omega^2 - \omega_0^2} \end{aligned}\right\}, \tag{728}$$

where

$$\omega_0 = \frac{1}{\sqrt{L_1C_1}} = \frac{1}{\sqrt{L_2C_2}} \tag{728a}$$

is the common frequency at which z_{1k} has its zero and z_{2k} its pole. The fact that these reactances are inverse with respect to R^2 gives

$$z_{1k}z_{2k} = \frac{L_1}{C_2} = \frac{L_2}{C_1} = R^2. \tag{729}$$

From the first relation (728) we then have

$$x_k = \frac{L_1}{2R} \cdot \frac{\omega^2 - \omega_0^2}{\omega}. \tag{730}$$

At the left-hand boundary this should equal -1, and at the right-hand boundary it should equal $+1$. Denoting these boundaries by ω_{c1} and ω_{c2}, respectively, we have

$$\left.\begin{aligned} \frac{L_1}{2R} \cdot \frac{\omega_{c1}^2 - \omega_0^2}{\omega_{c1}} &= -1 \\ \frac{L_1}{2R} \cdot \frac{\omega_{c2}^2 - \omega_0^2}{\omega_{c2}} &= 1 \end{aligned}\right\}. \tag{731}$$

From these we have

$$\frac{\omega_{c1}^2 - \omega_0^2}{\omega_{c1}} = -\frac{\omega_{c2}^2 - \omega_0^2}{\omega_{c2}},$$

which gives

$$\omega_{c1}\,\omega_{c2} = \omega_0^2. \tag{732}$$

Thus ω_0 is the *geometric mean* between the boundary frequencies. Substituting this back into either of (731) and writing

$$w = \omega_{c2} - \omega_{c1} \tag{733}$$

for the *width* of the pass band in radians per second, we find

$$\frac{L_1 w}{2R} = 1, \tag{734}$$

so that (730) becomes

$$x_k = \frac{\omega_0}{w}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right). \tag{730a}$$

For the determination of the parameters we use (728a), (729), and (734) and have

$$\left.\begin{aligned} L_1 &= \frac{2R}{w}; \quad C_1 = \frac{w}{2R\omega_0^2} \\ L_2 &= \frac{Rw}{2\omega_0^2}; \quad C_2 = \frac{2}{Rw} \end{aligned}\right\}. \tag{735}$$

It is interesting to note here that L_1 and C_2 are independent of the geometric mean frequency ω_0. The expressions for these parameters are very similar to (719) for the **L.P.** case, the only difference being that w takes the place of ω_c. The derived expression

$$w = \frac{2}{\sqrt{L_1 C_2}} \tag{736}$$

is identical in form to (720) for ω_c.

Considering the **B.E.** *filter* we recognize that z_{1k} must have a zero at the origin and at infinity. It thus becomes an anti-resonant component, so that z_{2k} must be a resonant component. As compared to the **B.P.** case the reactances are interchanged. We, therefore, have in place of (728)

$$\left.\begin{aligned} z_{1k} &= \frac{jL_1\omega}{1 - L_1C_1\omega^2} = \frac{\omega}{jC_1}\cdot\frac{1}{\omega^2 - \omega_0^2} \\ z_{2k} &= \frac{1 - L_2C_2\omega^2}{jC_2\omega} = \frac{jL_2}{\omega}\cdot(\omega^2 - \omega_0^2) \end{aligned}\right\}, \tag{737}$$

where ω_0 is now the frequency at which z_{1k} has its pole and z_{2k} its zero, i.e.,

$$\omega_0 = \frac{1}{\sqrt{L_1C_1}} = \frac{1}{\sqrt{L_2C_2}}, \tag{737a}$$

which is the same as (728a). The defining equation for the constant-k type gives

$$z_{1k}\, z_{2k} = \frac{L_1}{C_2} = \frac{L_2}{C_1} = R^2, \tag{738}$$

the same as (729) for the **B.P.** case.

From the first relation (737) we have

$$x_k = \frac{-\omega}{2RC_1\,(\omega^2 - \omega_0^2)}. \tag{739}$$

Denoting the lower and upper boundary frequencies between transmission and attenuation ranges by ω_{c1} and ω_{c2}, respectively, we note that in this case ω_{c1} becomes a right-hand boundary for the transmission range and ω_{c2} a left-hand boundary. Hence we have

$$\left.\begin{aligned} \frac{\omega_{c1}}{2RC_1(\omega_{c1}^2 - \omega_0^2)} &= -1 \\ \frac{\omega_{c2}}{2RC_1(\omega_{c2}^2 - \omega_0^2)} &= 1 \end{aligned}\right\}. \tag{740}$$

The simultaneous consideration of these equations again shows that ω_0 is the geometric mean between the boundaries ω_{c1} and ω_{c2}. Defining w as the width of the suppression band, we get from (740)

$$2RC_1w = 1, \tag{741}$$

so that (739) becomes

$$x_k = \frac{-1}{\dfrac{\omega_0}{w}\left(\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}\right)}. \tag{739a}$$

Comparing this with (730a) we see that for the same cut-off frequencies

$$x_{k\text{B.E.}} = \frac{-1}{x_{k\text{B P.}}}. \tag{742}$$

Note the similarity between this and (725).

Using (738) and (741) we find for the parameters

$$\left.\begin{aligned} C_1 &= \frac{1}{2Rw}; \; L_1 = \frac{2Rw}{\omega_0^2} \\ C_2 &= \frac{2w}{R\omega_0^2}; \; L_2 = \frac{R}{2w} \end{aligned}\right\}. \tag{743}$$

Here C_1 and L_2 are independent of ω_0. The expressions for these parameters are similar in form to those for the **H.P.** filter as given by (726). For the width of the suppression band we have

$$w = \frac{1}{2\sqrt{L_2C_1}}, \tag{744}$$

which is similar to (727).

In both the **B.P.** and **B.E.** cases it may be useful to express the boundary frequencies directly in terms of the band width w and the geometric mean frequency ω_0. Considering (732) and (733) simultaneously we find

$$\left.\begin{aligned} \omega_{c2} &= \sqrt{\omega_0^2 + \left(\frac{w}{2}\right)^2} + \frac{w}{2} \\ \omega_{c1} &= \sqrt{\omega_0^2 + \left(\frac{w}{2}\right)^2} - \frac{w}{2} \end{aligned}\right\}. \tag{745}$$

The calculation and plotting of the functions γ_1 and γ_2 for the last two cases is most conveniently carried out by means of the expressions (730a) and (739a) for x_k together with the plots of Fig. 110. To facilitate this, plots may also be made for x_k. These are illustrated in Fig. 117. It should be noted that they show geometric symmetry throughout with respect to the point $\omega/\omega_0 = 1$. This is seen analytically from the expressions (730a) and (739a) which show that

$$x_k\left(\frac{\omega}{\omega_0}\right) = -x_k\left(\frac{\omega_0}{\omega}\right),$$

where the quantities in the parentheses are considered independent variables. This geometric symmetry is inherent in the reactance function z_{1k} for these cases.

As a consequence of the geometric symmetry of x_k versus frequency and the fact that γ_1 and γ_2 show arithmetic symmetry with regard to

x_k (as illustrated by Fig. 110), the attenuation and phase functions versus frequency show geometric symmetry with respect to the point $\omega = \omega_0$ for the **B.P.** and **B.E.** cases. These functions are illustrated in Fig. 118 for the **B.P.** case. On account of the symmetry, only one branch for γ_1 and γ_2 need be calculated. This simplifies the drawing of these curves.

In **B.P.** cases where

$$\frac{\omega_0}{w} \gg 1,$$

i.e., where the band width is narrow compared with the geometric mean frequency, Fig. 117 shows that x_k becomes almost a linear function of

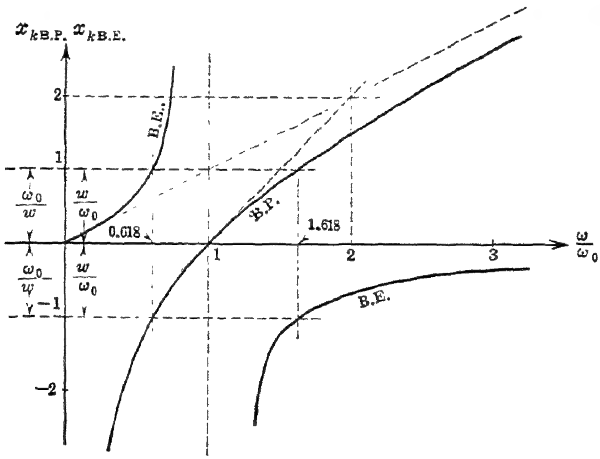


Fig. 117.—The frequency function (684) for the band-pass and band-elimination classes.

frequency over a considerable range in the vicinity of the pass band. For example, if $\omega_0 = 20\,w$, then the **B.P.** curve deviates very little from the tangent at $\omega = \omega_0$ for a range $-10 < x_k < 10$. In such a case the geometric symmetry of Fig. 118 will deviate very little from arithmetic symmetry in the vicinity of the pass band. It may also be mentioned in this connection that the geometric symmetry can be converted into arithmetic symmetry by using a logarithmic frequency scale. This is frequently done in work of this kind in order to be able to show the behavior over a wide frequency range on a reasonable length of paper. While this process sometimes has its advantages, it should be clearly recognized that it distorts the true state of affairs with regard to frequency dependence. The same comment applies, to a certain extent,

to the attenuation scale also, since this represents the logarithmic ratio of magnitudes except for reflection effects. The logarithmic function has the greatest power of submerging irregularities in the true functional behavior. The use of logarithmic scales, therefore, has the effect of making an otherwise poor behavior look quite tolerable. The student should not make the error of allowing himself to be misled by this artifice in gaining the true perspective in a given situation.

The mid-series and mid-shunt characteristic impedances for the **B.P.** and **B.E.** cases may be plotted versus frequency with the use of the plots of Fig. 111, which apply to all classes of the constant-k type, and the relations (730a) and (739a) or their graphs as given by Fig. 117. The same remarks regarding geometric symmetry apply to these

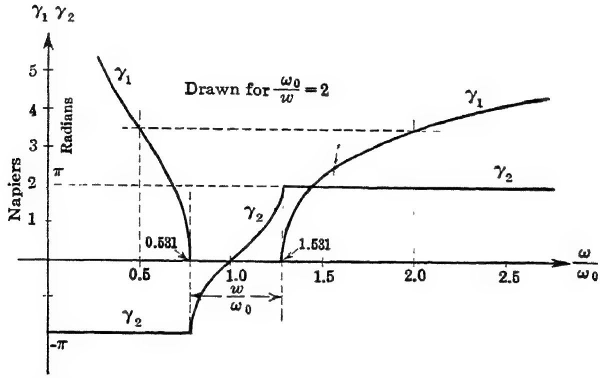


FIG. 118.—Attenuation and phase functions for the constant-k band-pass filter.

functions. The student may draw the curves for W_{1k} and W_{2k} corresponding to those of Fig. 118 as an exercise.

6. Shortcomings of the constant-k type. Although this type of filter structure may be designed to exhibit any desired allocation of transmission and attenuation ranges in the frequency spectrum, it is lacking in two major particulars. First, the characteristic impedance does not approximate a constant value very well over the transmission ranges. Second, the attenuation function does not rise very rapidly beyond the cut-off points; i.e., the boundaries are not as sharp or abrupt as we would like to have them. This latter point may be improved, of course, by the use of a larger number of sections in cascade, but the process is quite expensive and does not in the least improve the characteristic impedance.

In seeking to overcome these difficulties, Zobel[1] did not discard the

[1] See reference 2, p. 305.

constant-k type as a basic structure entirely, but attempted to improve its characteristics by adding other series and shunt branches to its terminations in a symmetrical fashion. The reactances in these additional branches are not chosen arbitrarily but must be derived from z_{1k} and z_{2k} in such a manner that the junction with the constant-k section presents no impedance irregularity, while the desired overall improvement is gained. The additional branches thus found form a selective structure by themselves which is also of the ladder type but which does not possess the property that the product of its component reactances equals a constant. Since in its derivation a new parameter m is introduced, Zobel calls it the **m-derived type.** It is not a type capable of functioning by itself with very desirable characteristics, but is designed to form part of a complete structure whose basic properties are provided by the constant-k type. The latter, which thus forms the nucleus for the resulting structure, is, in this capacity, referred to as the **prototype,** while the complete filter is called a **composite type.**

The process of deriving the reactances in the additional branches is difficult to present in logical form, and leads one to suspect that their original derivation was accomplished more by accident than by logic. This view is strengthened by the fact that the ultimate improvement in the characteristic impedance function is obtained only after the derived structure is segregated and properly associated with its prototype. This whole process may be justified, of course, on a heuristic basis, as, in fact, almost every significant step in synthesis problems must be. To the average student, however, a presentation which proceeds almost wholly on a heuristic basis fails to give complete satisfaction because it seems to arrive at results as if by accident rather than by logical anticipation. In the following presentation of this subject we have departed somewhat from Zobel's own form in an attempt partially to meet this situation.

7. The m-derived type. The approach to this problem may be stated in the following way. If we wish to derive a ladder structure to work in conjunction with the constant-k filter, it must possess two primary characteristics. First, it must have the same allocation of transmission and attenuation regions; and, second, either its mid-series or mid-shunt characteristic impedances must equal one of the constant-k characteristic impedances. The first requirement is obviously necessary. The second is necessary in order that the derived structure may be cascaded with its prototype without introducing an impedance irregularity at the junction. In order to effect an improvement, the derived structure should have an attenuation function which rises more abruptly at the boundaries of the transmission ranges. In addi-

tion to this, that characteristic impedance which is not equal to one of the constant-k characteristic impedances should show an improved behavior.

In order to see how such a derivation might be accomplished, the characteristic impedance and propagation functions may be expressed in the forms given by equations (424), (425), (427), and (428), pages 178–179. If we denote the reciprocals of the constant-k reactances by

$$y_{1k} = \frac{1}{z_{1k}};\ y_{2k} = \frac{1}{z_{2k}}, \tag{746}$$

we have

$$\gamma_k = \ln\left\{\frac{\sqrt{\dfrac{z_{1k}/2 + 2z_{2k}}{z_{1k}/2}} + 1}{\sqrt{\dfrac{z_{1k}/2 + 2z_{2k}}{z_{1k}/2}} - 1}\right\} = \ln\left\{\frac{\sqrt{\dfrac{y_{2k}/2 + 2y_{1k}}{y_{2k}/2}} + 1}{\sqrt{\dfrac{y_{2k}/2 + 2y_{1k}}{y_{2k}/2}} - 1}\right\}, \tag{747}$$

and

$$\left.\begin{aligned} W_{1k} &= \sqrt{\left(\frac{z_{1k}}{2}\right)\left(\frac{z_{1k}}{2} + 2z_{2k}\right)} \\ \frac{1}{W_{2k}} &= \sqrt{\left(\frac{y_{2k}}{2}\right)\left(\frac{y_{2k}}{2} + 2y_{1k}\right)} \end{aligned}\right\}, \tag{748}$$

where the subscript k has been added to γ to indicate that it refers to the constant-k type. These forms make it apparent that for the mid-series section the characteristic impedance and propagation functions depend upon the product and quotient, respectively, of the reactances

$$\left(\frac{z_{1k}}{2}\right) \quad \text{and} \quad \left(\frac{z_{1k}}{2} + 2z_{2k}\right),$$

whereas for the mid-shunt section they depend upon the product and quotient of the susceptances

$$\left(\frac{y_{2k}}{2}\right) \quad \text{and} \quad \left(\frac{y_{2k}}{2} + 2y_{1k}\right).$$

For brevity let us write

$$y_k = \sqrt{\frac{z_{1k}/2 + 2z_{2k}}{z_{1k}/2}} = \sqrt{\frac{y_{2k}/2 + 2y_{1k}}{y_{2k}/2}}, \tag{749}$$

so that

$$\gamma_k \doteq \ln\left(\frac{y_k + 1}{y_k - 1}\right). \tag{747a}$$

Then if we denote the series and shunt reactances of the derived structure by z_{1km}' and z_{2km}', respectively, and let

and
$$\left.\begin{aligned}\left(\frac{z_{1km}'}{2}\right) &= m\left(\frac{z_{1k}}{2}\right)\\ \left(\frac{z_{1km}'}{2} + 2z_{2km}'\right) &= \frac{1}{m}\left(\frac{z_{1k}}{2} + 2z_{2k}\right)\end{aligned}\right\}, \tag{750}$$

we will have for the mid-series characteristic impedance of the derived structure

$$\sqrt{\left(\frac{z_{1km}'}{2}\right)\left(\frac{z_{1km}'}{2} + 2z_{2km}'\right)} = W_{1k}, \tag{751}$$

while for the propagation function of the derived structure we find

$$\gamma_{km} = \ln\left(\frac{y_k + m}{y_k - m}\right). \tag{752}$$

The derivation indicated by (750) thus leaves the mid-series characteristic impedance invariant. It is, therefore, referred to as a **mid-series** or simply as a **series derivation.** The propagation function, on the other hand, is changed from the form (747*a*) to (752). The full significance of this will be appreciated from the more detailed discussion below. Meanwhile we note that, whereas the attenuation of the constant-k filter, according to (747*a*), becomes infinite for $y_k = 1$, that for the series derived type becomes infinite for $y_k = m$. Since the point at which infinite attenuation occurs obviously affects the rate at which the attenuation rises beyond any cut-off frequency, it is very likely that the abruptness of this rise may be controlled by means of the factor m.

From the relations (750) we may easily solve for the component reactances of the series derived type. We find

and
$$\left.\begin{aligned}z_{1km}' &= mz_{1k}\\ z_{2km}' &= \frac{z_{2k}}{m} + \frac{1 - m^2}{4m}z_{1k}\end{aligned}\right\}. \tag{753}$$

Thus the series reactance of the new structure is simply a factor times the series reactance of the constant k, while the new shunt reactance is a series combination of portions of the constant-k shunt and series reactances. Physical realizability in this simple form demands for the new *series* reactance that the factor m be real and positive, and

for the new *shunt* reactance that the factor $(1 - m^2)$ be positive. This restricts m to a real positive value between zero and unity,[1] i.e.,

$$0 < m \leq 1. \tag{754}$$

The limit $m = 1$ marks the trivial case for which the derived structure is identical to its prototype.

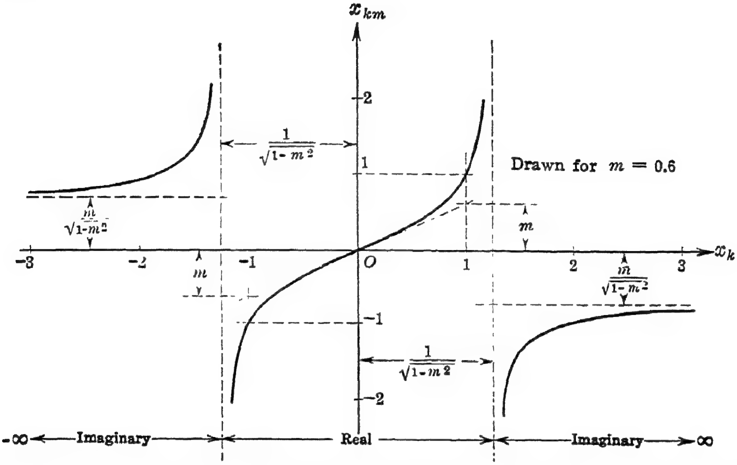


FIG. 119.—The frequency function (755) for the mid-series derived filter in terms of the corresponding function for the constant-k prototype according to (755a).

For the further investigation of the properties of the derived structure it is useful to determine the corresponding parameter x according to the substitution (678). In the notation of this type we have

$$-x_{km}^2 = \frac{z_{1km}'}{4z_{2km}'}. \tag{755}$$

[1] It will be seen shortly that this restriction on m leads to a rather poor phase characteristic in the pass band, i.e., one which is even more concave than that of the constant-k type. A more linear, or even convex, phase characteristic is obtained with m-values greater than unity. In fact, Bode has shown that, by admitting complex m-values, a more desirable behavior for the composite filter may be expected. Values of m larger than unity, or complex m-values, cannot in general be physically realized with structures of the form discussed here, but are realizable as *lattices*. An exception is found, for example, in the low-pass case where, by means of mutual coupling between the series inductances of a T-section, the negative value of $\left(\frac{1 - m^2}{4m}\right) \cdot z_{1k}$ in z_{2km}' may be realized for an m-value greater than unity. This method is quite commonly utilized to produce a fairly linear phase shift or delay network with good economy. With $m = 1.4$, linearity within a tolerance of 3.75 per cent is attained over 77 per cent of the theoretical pass band.

Substituting from (753) and recalling (684), we find

$$x_{km} = \frac{m x_k}{\sqrt{1 - (1 - m^2) x_k^2}}. \tag{755a}$$

A plot of x_{km} versus x_k is given in Fig. 119. Of primary importance here is the fact that the interval -1 to $+1$ for x_k corresponds to the same interval for x_{km}. Hence the allocation of transmission regions remains unchanged. This fact is also evident from a comparison of (747*a*) and (752) if we note that by (749)

$$y_k = \sqrt{1 - x_k^{-2}}. \tag{749a}$$

Thus y_k is a pure imaginary quantity for the interval $-1 \leq x_k \leq 1$, so that the argument of the logarithm in (747*a*) becomes the quotient of a complex number and its negative conjugate. The magnitude of such a quotient is unity, and hence γ_{1k} (real part of γ_k) is zero. Since m is a real number, the same argument applies to (752) for the derived type, so that the transmission regions for the latter and its prototype must be coincident.

Fig. 119 further shows that x_{km} becomes infinite at the frequencies corresponding to

$$x_k = \pm \frac{1}{\sqrt{1 - m^2}}. \tag{756}$$

By choosing m according to the range (754), this point may be made to fall anywhere between $x_k = \pm 1$ for $m = 0$, and $x_k = \infty$ for $m = 1$. Since the attenuation of the derived type becomes infinite where x_{km} becomes infinite, we see that this point of infinite attenuation can be made to fall anywhere between the cut-off point ($x_k = \pm 1$) and that point at which the constant-k filter has an infinite attenuation ($x_k = \infty$). The sharpness with which the attenuation function rises, therefore, is adjustable by the choice of m.

The specific determination of the functions γ_{1km} and γ_{2km} (real and imaginary parts of the propagation function for the derived type) may be accomplished by returning to the expressions (678*a*) and (679) which hold generally for the ladder structure. For the derived type x is x_{km}, given by (755*a*) in terms of x_k for the constant k. In applying (679) we note that x_{km} is *real* for the interval

$$\frac{-1}{\sqrt{1 - m^2}} \leq x_k \leq \frac{1}{\sqrt{1 - m^2}}, \tag{757}$$

and imaginary for

$$\frac{1}{\sqrt{1 - m^2}} \leq |x_k| < \infty. \tag{758}$$

For the interval (757) we have, therefore,

$$\sinh\frac{\gamma_{1km}}{2}\cdot\cos\frac{\gamma_{2km}}{2}=0.$$

This interval must be further subdivided into $-1\le x_k\le 1$ for which

$$\left.\begin{aligned}\gamma_{1km}&=0\\ \gamma_{2km}&=2\sin^{-1}x_{km}=2\tan^{-1}\frac{mx_k}{\sqrt{1-x_k^2}}\end{aligned}\right\},\tag{759}$$

and $1\le|x_k|\le(1-m^2)^{-1/2}$ for which

$$\left.\begin{aligned}\gamma_{1km}&=2\cosh^{-1}|x_{km}|=\ln\left(\frac{m|x_k|+\sqrt{x_k^2-1}}{m|x_k|-\sqrt{x_k^2-1}}\right)\\ \gamma_{2km}&=\pm\pi\end{aligned}\right\}.\tag{760}$$

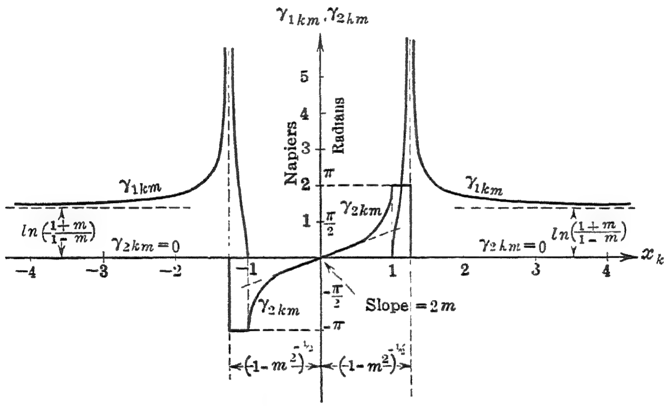


FIG. 120.—Attenuation and phase functions for the derived type in terms of the frequency function (684) of the constant-k prototype.

Here the plus or minus sign for γ_{2km} is chosen according to whether x_k is positive or negative, respectively. This is again justified on the basis that the slope of γ_{2km} versus x_{km} must be positive.

For the interval (758) we have by (679)

$$\cosh\frac{\gamma_{1km}}{2}\cdot\sin\frac{\gamma_{2km}}{2}=0,$$

which is satisfied only by

$$\left.\begin{aligned}&\gamma_{2km}=0\\ \text{whence}\qquad &\gamma_{1km}=2\sinh^{-1}|x_{km}|=\ln\left(\frac{\sqrt{x_k^2-1}+m|x_k|}{\sqrt{x_k^2-1}-m|x_k|}\right)\end{aligned}\right\}.\tag{761}$$

Plots of these functions versus x_k are shown in Fig. 120. Since the

function x_k versus frequency is given by the corresponding constant-k design, this figure applies generally to the derived type of any class.

It is interesting to compare these curves with those of Fig. 110 for the constant-k type. The derived type has much sharper boundaries owing to the fact that the points of infinite attenuation lie near the values $x_k = \pm 1$ instead of at $x_k = \pm \infty$. Beyond these points of infinite attenuation, however, the attenuation characteristic drops off quite rapidly and approaches the asymptotic value

$$\ln\left(\frac{1+m}{1-m}\right). \tag{762}$$

This value decreases with m and becomes zero for $m = 0$. Since the sharpness of cut-off increases with a decrease in m, we see that the

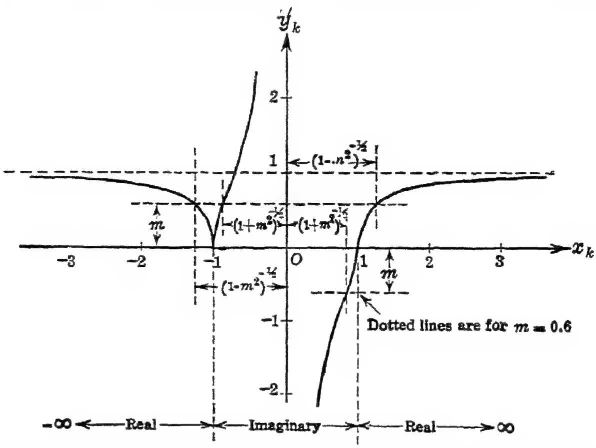


FIG. 121.—Plot of the function defined by eqs. (749) and (749a) versus the frequency function (684) of the constant-k filter.

boundaries cannot be made abrupt without sacrificing attenuation in the regions beyond the points of infinite attenuation. However, since the derived type is to operate in cascade with the constant-k, the latter's attenuation characteristic will assume the attenuation burden where the derived type fails. The details of this will be discussed later.

The phase characteristic of the derived type differs from that for the constant-k, in two particulars. First, its slope at the origin is only m times as large; and, second, it drops to zero beyond the points of infinite attenuation.

An alternative method of determining the attenuation and phase functions of the derived type is by means of the relations (752) and (749a). A plot of y_k versus x_k according to (749a) is shown in Fig. 121.

In the transmission region y_k is imaginary; in the attenuation region it is real. Where y_k is imaginary and equal to m in magnitude, the phase function equals $\pm \pi/2$. Where y_k is real and equal to m, the attenuation is infinite.

We have yet to investigate the mid-shunt characteristic impedance of the mid-series derived type. This is the characteristic impedance of a Π-section with the component series and shunt reactances z_{1km}' and z_{2km}'. By (424) and (427a), pages 178–179, we note that for the uniform ladder structure

$$\frac{Z_T}{Z_\pi} = \left(1 + \frac{z_1}{4z_2}\right). \tag{763}$$

Denoting the mid-shunt characteristic impedance in this case by W_{2km}, we have with (755) and (755a)

$$\frac{W_{1k}}{W_{2km}} = (1 - x_{km}^2) = \frac{1 - x_k^2}{1 - (1 - m^2)x_k^2}, \tag{763a}$$

so that

$$W_{2km} = \frac{W_{1k}\{1 - (1 - m^2)\,x_k^2\}}{1 - x_k^2}, \tag{764}$$

and with (688) we find

$$W_{2km} = \frac{R\{1 - (1 - m^2)x_k^2\}}{\sqrt{1 - x_k^2}}, \tag{764a}$$

or

$$W_{2km} = W_{2k}\{1 - (1 - m^2)x_k^2\}. \tag{764b}$$

Thus the mid-series derived type has a mid-shunt characteristic impedance which is a function of a new form. A plot of (764a) versus x_k is shown in Fig. 122. This should be compared to the plot of W_{2k} shown in Fig. 111. We see that W_{2km} approximates the nominal value R much better over the transmission region. Like W_{1k} it is real in the transmission region and imaginary in the attenuation ranges. In the latter it has a positive slope like a physical reactance.

The manner in which W_{2km} behaves in the transmission range depends upon the parameter m. In this range the function has a *maximum* value equal to R at $x_k = 0$, and *minima* at

$$x_k = \pm\sqrt{\frac{1 - 2m^2}{1 - m^2}} \tag{765}$$

equal to

$$(W_{2km})_{\min} = 2Rm\sqrt{1 - m^2}. \tag{765a}$$

It again passes through the value R at the points

$$x_k = \pm \frac{\sqrt{1 - 2m^2}}{1 - m^2}. \tag{765b}$$

The minima according to (765) exist only when the quantity under the radical is positive, i.e., for

$$m < \frac{1}{\sqrt{2}} = 0.707. \tag{764b}$$

When m equals this limiting value the minima merge with the maximum at $x_k = 0$. For larger values of m the function merely has a minimum

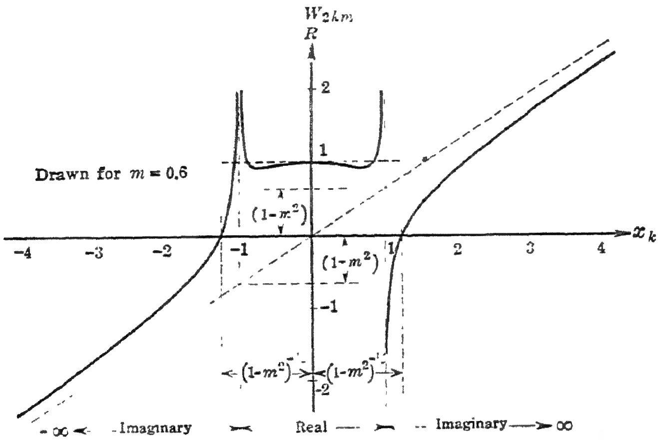


FIG. 122.—Mid-shunt characteristic impedance of the series derived type plotted versus the frequency function (684) of the constant-k filter.

at $x_k = 0$. In the limit $m = 1$, W_{2km} equals W_{2k}, as may be seen from (764b). In the other limiting case $m = 0$, W_{2km} becomes equal to

$$R\sqrt{1 - x_k^2}$$

which is W_{1k}. Hence we see that, as m passes from zero to unity, W_{2km} passes from W_{1k} to W_{2k} in form.

The behavior of W_{2km} in the transmission region, so far as its ability to approximate its nominal value is concerned, is best expressed in terms of a maximum tolerance and coverage. To illustrate this, W_{2km} is plotted to a larger scale for the region $0 \leq x_k \leq 1$ in Fig. 123. In order to see how the coverage and tolerance depend upon m, these may be calculated for a number of assumed m-values and the results plotted

versus m. This is done in Fig. 124 where, instead of the coverage, the percentage not covered is plotted in order that the same scale may be

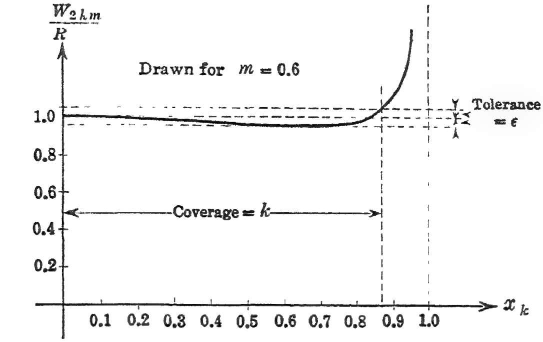


FIG. 123.—The impedance of Fig. 122 plotted in the right-half of the transmission range to a larger scale.

used as for the tolerance. Since tolerances above about 25 per cent are impracticable, the curves are not carried beyond this value. The tolerance is expressed with the help of (765a) as

$$\epsilon = \frac{(W_{2km})_{\max} - (W_{2km})_{\min}}{(W_{2km})_{\max}} = 1 - 2m\sqrt{1 - m^2}, \tag{766}$$

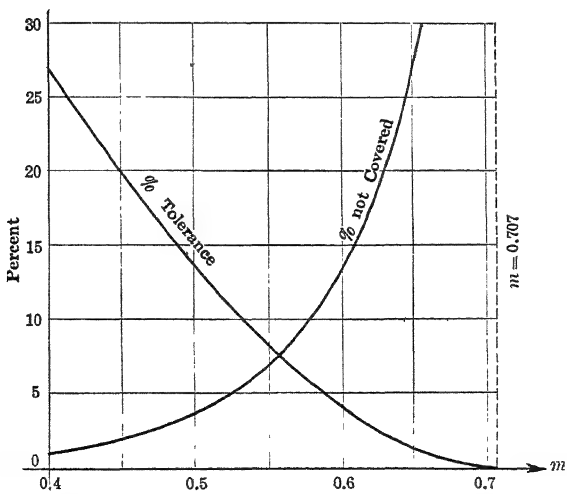


FIG. 124.—Tolerance and coverage of the function of Fig. 123 in terms of the parameter m with the resistance R as a norm.

and the coverage as the positive real root of the equation

$$\frac{1 - (1 - m^2)x_k^2}{\sqrt{1 - x_k^2}} = 1 + \epsilon. \tag{767}$$

We see that if the filter is terminated in a mid-shunt end of the series derived type the possibilities of having its characteristic impedance approximate a constant resistance are very much better than for the constant-k type alone. Usually the terminal resistance is made equal to R, as has been assumed in the determination of tolerance and coverage above. This is not necessary, however. In fact, if the terminal

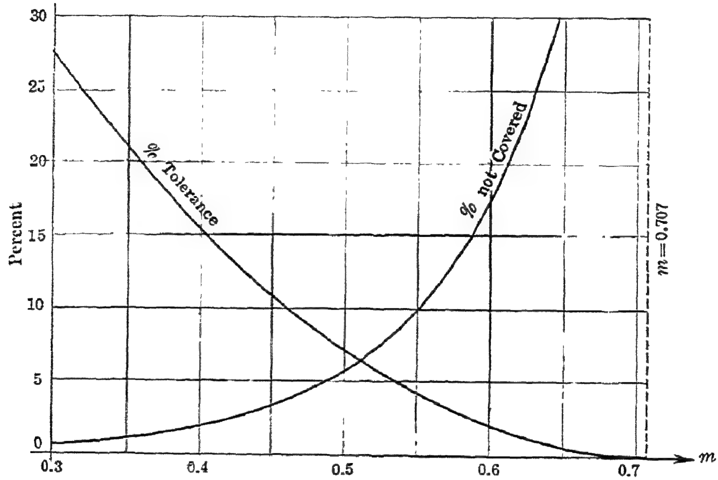


FIG. 125.—Tolerance and coverage of the function of Fig. 123 in terms of the parameter m with the value (768) or (768a) as a norm.

resistance is made somewhat smaller than this value, better combinations of tolerance and coverage will result. For this basis we make the terminal resistance equal to

$$Z_S = Z_R = \frac{(W_{2km})_{\max} + (W_{2km})_{\min}}{2}. \tag{768}$$

The tolerance is then defined as

$$\epsilon' = \frac{(W_{2km})_{\max} - (W_{2km})_{\min}}{(W_{2km})_{\max} + (W_{2km})_{\min}}, \tag{769}$$

and the coverage as the positive real non-vanishing root of

$$W_{2km} = (W_{2km})_{\max}. \tag{770}$$

Substituting from above this gives

$$Z_S = Z_R = R(m\sqrt{1 - m^2} + \tfrac{1}{2}), \tag{768a}$$

$$\epsilon' = \frac{1 - 2m\sqrt{1 - m^2}}{1 + 2m\sqrt{1 - m^2}}, \tag{769a}$$

and for the coverage

$$k' = \frac{\sqrt{1 - 2m^2}}{1 - m^2}. \tag{770a}$$

Plots of tolerance and per cent not covered for this basis are shown in Fig. 125.

On the basis of Figs. 123 and 124 the terminal resistance equals the nominal resistance R and hence equals the characteristic impedance at the point $x_k = 0$, while on the basis of Fig. 125 the terminal resistance is fixed by (768a) and thus the ability of W_{2km} to approximate a constant is more fully utilized. The point $x_k = 0$ in this case becomes a point of maximum deviation. This process is similar to terminating the constant-k filter in a resistance other than the nominal value so that a better average agreement is obtained.

The practice indicated by Zobel utilizes the function W_{2km} according to Figs. 123 and 124, however. Here the parameter m is usually chosen equal to 0.6 which gives rise to a tolerance of 4 per cent over about 86 per cent of the band. The Figs. 119 to 123 are all drawn for this value of m. The reader should note that this fixes the points of infinite attenuation at $x_k = \pm 1.25$. A sharper cut-off than this requires a smaller value for m; but this increases the tolerance and decreases the attenuation in the interior of the attenuation region. Hence we see that the choice of m involves a compromise.

The mid-series derivation indicated by (750), leading to the new component reactances (753), therefore, fulfils the expectations mentioned earlier. Namely, it gives rise to a structure whose mid-series termination may be joined to that of the constant-k without impedance irregularity; it leads to sharper cut-off properties; and, finally, it makes available a new characteristic impedance in the function W_{2km} which possesses marked superiority in its ability to approximate a constant value over the transmission range.

Substantially the same results may also be obtained by proceeding in a manner whereby the mid-shunt characteristic impedance remains invariant. This process, which leads to the so-called **mid-shunt derived type**, may more conveniently be carried out in terms of susceptances instead of reactances. Noting the second relation (748), we see that a new structure with component susceptances defined by

$$\left.\begin{aligned} \left(\frac{y_{2km}''}{2}\right) &= m\left(\frac{y_{2k}}{2}\right) \\ \left(\frac{y_{2km}''}{2} + 2y_{1km}''\right) &= \frac{1}{m}\left(\frac{y_{2k}}{2} + 2y_{1k}\right) \end{aligned}\right\} \tag{771}$$

will have a mid-shunt characteristic impedance equal to W_{2k}, while for its propagation function, according to (747) and (749), it will have

$$\gamma_{km} = \ln\left(\frac{y_k + m}{y_k - m}\right). \tag{752}$$

which is the same as for the series derivation. By (771) this structure has the component susceptances

$$\left.\begin{aligned} y_{2km}'' &= m y_{2k} \\ y_{1km}'' &= \frac{y_{1k}}{m} + \frac{1 - m^2}{4m} y_{2k} \end{aligned}\right\}. \tag{772}$$

Here the physical realization of these susceptances in the form of simple combinations of those for the constant-k filter again demands that m be a real positive number between zero and unity.

The corresponding reactances are the reciprocals of these, namely

$$\left.\begin{aligned} z_{2km}'' &= \frac{z_{2k}}{m} \\ z_{1km}'' &= \cfrac{1}{\cfrac{1}{m z_{1k}} + \cfrac{1}{\cfrac{4m}{1 - m^2} z_{2k}}} \end{aligned}\right\}. \tag{772a}$$

Hence the new shunt reactance is a factor times the constant-k shunt reactance, and the new series reactance is a parallel combination of portions of the series and shunt reactances of the constant k.

Since the propagation function of the shunt derived type is the same as that of the series derived type, the ratio of component reactances in the two must be the same. This follows from (755) with the fact that the function x_{km} determines the propagation function. Thus we have

$$\frac{z_{1km}'}{z_{2km}'} = \frac{z_{1km}''}{z_{2km}''}, \tag{773}$$

and, by the first equations (753) and (772a), this gives

$$z_{1km}' z_{2km}'' = z_{1km}'' z_{2km}' = z_{1k} z_{2k} = R^2. \tag{773a}$$

Hence the series reactance of the series derivation is the inverse with respect to R^2 of the shunt reactance of the shunt derivation, and likewise for the series and shunt reactances of the shunt and series derivations, respectively. Thus the reactances of one derivation may be obtained from those of the other by reciprocation with respect to R^2. The propagation function of the shunt derived type needs no further comment since it is the same as that for the series derivation.

The shunt derived structure has a mid-series characteristic impedance which we shall denote by W_{1km}. The relation (763a) may then be extended so as to read

$$\frac{W_{1k}}{W_{2km}} = \frac{W_{1km}}{W_{2k}} = 1 - x_{km}^2 = \frac{1 - x_k^2}{1 - (1 - m^2)x_k^2}. \tag{774}$$

From this we get the interesting relation

$$W_{1km} \cdot W_{2km} = W_{1k} \cdot W_{2k} = R^2, \tag{775}$$

i.e., the mid-series impedance of the shunt derivation and the mid-shunt impedance of the series derivation are inverse functions with respect to R^2. Using (764a) this gives

$$\left.\begin{aligned} W_{1km} &= \frac{R\sqrt{1 - x_k^2}}{1 - (1 - m^2)x_k^2} \\ &= \frac{W_{1k}}{1 - (1 - m^2)x_k^2} \end{aligned}\right\}. \tag{774a}$$

The form of W_{1km} versus x_k may be obtained by reciprocating Fig. 122 with respect to R^2. The same applies to Fig. 123 for the transmission region alone. The student may sketch these as an exercise. Regarding the dependence of tolerance and coverage upon the parameter m, this will be slightly altered when expressed according to the basis of Figs. 123 and 124. The reason for this is that W_{1km} has a minimum value of R at $x_k = 0$, and maxima at the points (765) where W_{2km} has its minima. These maxima are, according to (765a) and (775),

$$(W_{1km})_{\max} = \frac{R}{2m\sqrt{1 - m^2}}, \tag{776}$$

so that the tolerance expressed in terms of the nominal value R becomes

$$\epsilon = \frac{(W_{1km})_{\max} - (W_{1km})_{\min}}{(W_{1km})_{\min}} = \frac{1}{2m\sqrt{1 - m^2}} - 1 \tag{777}$$

instead of (766), while the coverage becomes the positive real root of the equation

$$\frac{\sqrt{1 - x_k^2}}{1 - (1 - m^2)x_k^2} = 1 - \epsilon. \tag{778}$$

For small tolerances, however, the results are substantially the same as for W_{2km} on this basis. For example, a tolerance of 4 per cent for W_{2km} becomes a tolerance of 4.17 per cent for W_{1km}. Hence the Fig. 124 may be used for W_{1km} with sufficient accuracy for most practical purposes.

However, when the terminal resistance is made equal to

$$Z_S = Z_R = \frac{(W_{1km})_{\max} + (W_{1km})_{\min}}{2}, \tag{779}$$

and the tolerance expressed with respect to this value, namely

$$\epsilon' = \frac{(W_{1km})_{\max} - (W_{1km})_{\min}}{(W_{1km})_{\max} + (W_{1km})_{\min}}, \tag{780}$$

while the coverage is defined as the positive real non-vanishing root of

$$W_{1km} = (W_{1km})_{\min}, \tag{781}$$

then the expressions for tolerance and coverage for W_{1km} become the same as (769a) and (770a) for W_{2km}. This may be seen by substituting (775) into (780) and (781). They then reduce to (769) and (770),

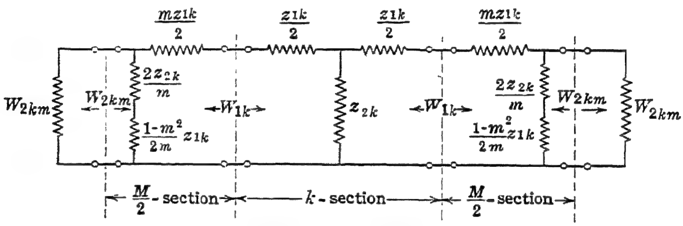


FIG. 126.—Composite filter consisting of a constant-k T-section with symmetrically placed M-type half-sections.

respectively. Hence Fig. 125 holds for W_{1km} as well as for W_{2km}. The proper terminal resistances are different, however. Namely, (779) gives

$$Z_S = Z_R = \frac{R(1 + 2m\sqrt{1 - m^2})}{4m\sqrt{1 - m^2}} \tag{779a}$$

as compared with (768a). The difference is usually very small.

8. The composite filter. As mentioned above, the derived type is utilized to best advantage when cascaded with the constant-k. In order to gain not only the improved cut-off properties but also the advantages of closer matching over the transmission range, the derived type, either series or shunt, is not cascaded with the constant-k in the form of a whole section, but is bisected and the halves are placed at either end of the constant-k in a symmetrical fashion. In this way the terminations are mid-shunt when the mid-series derivation is used, and mid-series when the mid-shunt derivation is used. Thus the terminal resistances are matched to either W_{2km} or W_{1km}, respectively. This process involves the use of a symmetrical constant-k structure with mid-series or mid-shunt ends according to whether the derived

section is mid-series or mid-shunt. The resulting filter is called *composite*. Fig. 126 illustrates this for the case where mid-series derived half-sections are associated with a single T-section of constant-k.

Here the adjacent series arms may be combined into a single reactance of the value $\frac{1+m}{2} z_{1k}$. Thus the composite structure of Fig. 126 contains seven reactances, each of which is a factor times either z_{1k} or z_{2k}. Since the latter each contain as many elements as the filter has cut-off points, the composite structure in its simplest form involves $7k$ elements, where k equals the number of cut-off frequencies. This is also true for the constant-k Π-section plus mid-shunt derived half-sections. There the adjacent shunt arms may be combined into a single reactance.

According to (747a) and (752), the propagation function for the composite filter consisting of one constant-k section plus a pair of m-derived half-sections is given by

$$\gamma_k + \gamma_{km} = \ln\left(\frac{y_k + 1}{y_k - 1} \cdot \frac{y_k + m}{y_k - m}\right), \tag{782}$$

where y_k is expressed in terms of the component constant-k reactances by means of (749) or in terms of x_k by means of (749a). In particular, the attenuation function is given by

$$\gamma_{1k} + \gamma_{1km} = \ln\left|\frac{y_k + 1}{y_k - 1} \cdot \frac{y_k + m}{y_k - m}\right|. \tag{782a}$$

Since the constant-k section provides points of infinite attenuation at $x_k = \pm\infty$, and the derived section adds infinite attenuation points at $x_k = \pm(1 - m^2)^{-1/2}$, the resulting attenuation characteristic (782a) shows three infinite points for each internal attenuation band, one of these due to the constant-k, and two due to the m-derived. The latter two are located near the cut-off frequencies while the former lies between these. (For the **B.E.** filter this point is the geometric mean between the cut-offs.) Between these three points of infinite attenuation, the function (782a) has two minima. In a practical case it is necessary to evaluate these minima in order to determine whether the resulting attenuation is sufficient at these points. This we shall now do.

Since in an attenuation region the argument of the logarithm in (782a) is greater than unity, the minima of the attenuation are coincident with those for the argument. Hence we differentiate this argument with respect to y_k and set the result equal to zero. This gives

$$m(y_k^2 - 1) + y_k^2 - m^2 = 0$$

or

$$(y_k)_{\min} = \pm\sqrt{m}, \tag{783}$$

which, by (749a), is the same as

$$(x_l)_{\min} = \frac{\pm 1}{\sqrt{1-m}}. \tag{783a}$$

The minimum value of attenuation at these points is found by substituting (783) back into (782a). This gives

$$\left.\begin{aligned}(\gamma_{1k} + \gamma_{1km})_{\min} &= \ln\left(\frac{1+\sqrt{m}}{1-\sqrt{m}}\cdot\frac{1+\sqrt{m}}{1-\sqrt{m}}\right)\\ &= 2\ln\left(\frac{1+\sqrt{m}}{1-\sqrt{m}}\right)\end{aligned}\right\}. \tag{784}$$

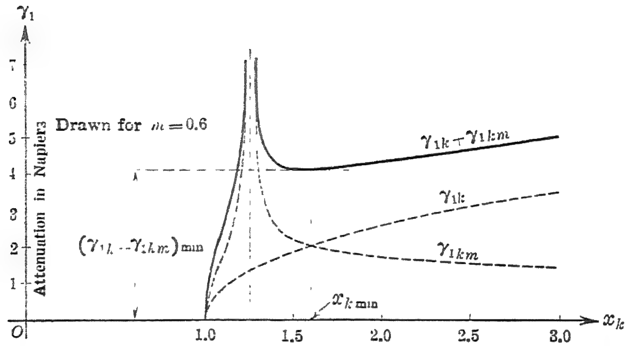


FIG. 127.—Attenuation of the component parts (dotted curves) and of the composite filter of Fig. 126 in terms of the frequency function (684) of the constant-k prototype.

This substitution incidentally shows that at the point (783a)

$$\gamma_{1k} = \gamma_{1km},$$

i.e., the minimum in the resultant attenuation occurs where the separate attenuation curves for the constant-k and the derived sections cross. The minimum value of attenuation is twice that for either section. Fig. 127 illustrates this for $m = 0.6$. In this case

$$\begin{aligned}(\gamma_{1k} + \gamma_{1km})_{\min} &= 2\ln 7.89 = 4.13 \text{ napiers}\\ &= 35.88 \text{ db},\end{aligned}$$

which is still a reasonably high attenuation. However, for $m = 0.3$ we would have

$$\begin{aligned}(\gamma_{1k} + \gamma_{1km})_{\min} &= 2\ln 3.43 = 2.46 \text{ napiers}\\ &= 21.4 \text{ db}.\end{aligned}$$

This is rather low. On the other hand, the rise beyond the cut-off in Fig. 127 is not as steep as might be desired. In this connection it must

be remembered that a smaller m-value also increases the tolerance in W_{1km} or W_{2km}. The usual compromise is to choose $m = 0.6$.

It is, of course, possible to use more than one constant-k section in the composite filter in order to increase the attenuation. It is also possible to use an odd number of constant-k half-sections with a mid-series derived half-section at one end and a mid-shunt derived half-section at the other. Here it must be remembered, however, that this causes the introduction of the dissymmetry factor

$$\sqrt{\frac{\mathfrak{A}}{\mathfrak{D}}} = \sqrt{\frac{W_{1km}}{W_{2km}}} = \frac{W_{1km}}{R} \tag{785}$$

in the voltage and current ratios, as given in equations (411) and (412), page 174. For the volt-ampere ratio this factor cancels out, as discussed earlier, so that the use of half-sections does not disturb matters when this ratio governs.

When more than one constant-k section is used, the minimum of the net attenuation no longer falls at the point (783) or (783a). For example, with two constant-k sections plus one m-derived we have the net attenuation

$$2\gamma_{1k} + \gamma_{1km} = \ln \left|\frac{y_k + 1}{y_k - 1}\right|^2 \cdot \left|\frac{y_k + m}{y_k - m}\right|. \tag{786}$$

Differentiating the argument and setting it equal to zero gives

$$\left.\begin{aligned} (y_k)_{\min} &= \pm\sqrt{\frac{m(2m+1)}{m+2}} \\ \text{or} \qquad (x_k)_{\min} &= \pm\sqrt{\frac{m+2}{2(1-m^2)}} \end{aligned}\right\}. \tag{787}$$

Substituting back into (786) we have for this case

$$(2\gamma_{1k} + \gamma_{1km})_{\min} = \ln\left(\frac{(2m+1)^{\frac{3}{2}} + m^{\frac{1}{2}}(m+2)^{\frac{3}{2}}}{(2m+1)^{\frac{3}{2}} - m^{\frac{1}{2}}(m+2)^{\frac{3}{2}}}\right). \tag{788}$$

9. Repeated derivations. Although the derived type offers a considerable improvement upon the constant-k when cascaded with this in the manner just described, the results obtainable are restricted owing to the fact that only one additional design parameter is introduced. This, of course, may be remedied in a way by cascading two m-sections with the constant-k, which have two different values of m. For example, an additional mid-series m-section may be introduced between the constant-k and one of the half-sections in the filter of Fig. 126. This additional m-section may be designed with a smaller value of m so

as to improve the sharpness of cut-off still more while the half-sections on the ends provide the more desirable form for W_{2km} with a value of m chosen for this purpose primarily.

This process lends itself readily to a variety of modifications and on the whole proves to be a flexible method of meeting varying practical requirements. The logical continuation of the composite filter structure of Fig. 126, however, is to add half-sections in a symmetrical fashion which bear the same relation to the derived type as the latter does to its prototype. In other words, the same process of derivation may be applied to the above derived type to obtain a so-called **double m-derived type.** In this way not only the attenuation function but also the characteristic impedance at the final terminals may be improved still more. This process, which we shall now discuss, may be repeated as often as desired, so that ultimately the most rigid specifications regarding sharpness of cut-off, minimum attenuation in the attenuation region, and tolerance and coverage for the characteristic impedance may be met.[1]

The fundamental idea involved here is that the series and shunt derivations indicated by (750) and (771), respectively, are not restricted to any particular type of structure except that it be of the uniform ladder form. Thus the component reactances or susceptances of either the mid-series or mid-shunt derived types can be substituted in place of those of the constant-k in these relations. The last have the property of keeping the mid-series or mid-shunt characteristic impedances invariant no matter what the component reactances are.

This means that we could proceed to find four additional derived types, two from the mid-series and two from the mid-shunt derived structures which we already have. A moment's reflection will show, however, that one out of each two of these additional derivations has no utility. Thus a second series derivation in terms of the reactances of the first series derivation, or a second shunt derivation in terms of the reactances of the first shunt derivation, would lead to structures which could not be utilized to best advantage. In order logically to carry on the process involved in the composite filter of Fig. 126, the terminations of this should be followed by half-sections which are shunt-derived from the reactances of the first series derivation. These in turn may be followed symmetrically by half-sections which are series-derived from the reactances of the preceding shunt derivation, and so on. In other words, the successively added half-sections in the

[1] This is discussed in a paper by O. J. Zobel, entitled "Extensions to the Theory and Design of Electric Wave-Filters," B.S.T.J., April, 1931, p. 284.

symmetrical fashion of Fig. 126 should be alternately series- and shunt-derived from the reactances of each preceding derivation in order that the composite structure at any stage may fully utilize the potential improvements in the propagation and characteristic impedance functions which each derivation offers.

Such a scheme may evidently take two different forms depending upon whether a T- or Π-section of constant-k is chosen as the "nucleus" upon which to build the composite structure. If the T-section is chosen, the process is said to follow a **series sequence**; if the Π-section is chosen, it follows a **shunt sequence.**[1] In each case the derivations are alternately series and shunt, of course, but the one starts with a series derivation in terms of the constant-k while the other starts with a shunt derivation. The end results obtainable with either sequence are the same for a given number of derivations. The propagation functions encountered at any stage of the two sequences are identical, while the characteristic impedance functions are inverse with respect to the resistance R^2.

In the series sequence the component reactances are denoted by

$$(z_{1k}, z_{2k});\ (z_{1km}', z_{2km}');\ (z_{1kmm'}'', z_{2kmm'}'');\ (z_{1kmm'm''}', z_{2kmm'm''}')\ ;\ \text{etc.},$$

while the characteristic impedances are denoted by

$$W_{1k} \rightarrow W_{1k},\ W_{2km} \rightarrow W_{2km},\ W_{1kmm'} \rightarrow W_{1kmm'},\ W_{2kmm'm''} \cdots \text{etc.}$$

In the shunt sequence the component reactances are denoted by

$$(z_{1k}, z_{2k});\ (z_{1km}'', z_{2km}'');\ (z_{1kmm'}', z_{2kmm'}');\ (z_{1kmm'm''}'', z_{2kmm'm''}'');\ \text{etc.},$$

while the characteristic impedances are denoted by

$$W_{2k} \rightarrow W_{2k},\ W_{1km} \rightarrow W_{1km},\ W_{2kmm'} \rightarrow W_{2kmm'},\ W_{1kmm'm''} \cdots \text{etc.}$$

A *single* prime on the reactances indicates a *series* derivation and a *double* prime a *shunt* derivation, while the successive derivations in either sequence employ the parameters m, m', m'', etc. These letters in the subscripts indicate the order of the derived quantity. Thus the quantities in either sequence with subscripts m are the result of the first derivation and are said to belong to the ***M*-type**; those with subscripts mm' are the result of a second derivation and belong to the ***MM'*-type**; those with subscripts $mm'm''$ belong to the ***MM'M''*-type**, etc.

Thus the component reactances of the ***MM'*-type** in the series sequence

[1] These are also referred to as *sequence 1* and *sequence 2*, respectively.

are the result of a shunt derivation from z_{1km}' and z_{2km}' according to the relations (772a), and hence are given by

$$\left.\begin{aligned} z_{2kmm'}'' &= \frac{z_{2km}'}{m'} \\ z_{1kmm'}'' &= \cfrac{1}{\cfrac{1}{m'z_{1km}'} + \cfrac{1}{\cfrac{4m'}{1-m'^2}z_{2km}'}} \end{aligned}\right\}. \tag{789}$$

Substituting from (753) this gives

$$\left.\begin{aligned} z_{2kmm'}'' &= \frac{z_{2k}}{mm'} + \frac{1-m^2}{4mm'}z_{1k} \\ z_{1kmm'}'' &= \cfrac{1}{\cfrac{1}{mm'z_{1k}} + \cfrac{1}{\cfrac{m'(1-m^2)}{m(1-m'^2)}z_{1k} + \cfrac{4m'}{m(1-m'^2)}z_{2k}}} \end{aligned}\right\}. \tag{789a}$$

These are simply series and parallel combinations of portions of z_{1k} and z_{2k}.

The component reactances of the *MM'*-type in the shunt sequence are the result of a series derivation from z_{1km}'' and z_{2km}'' according to the relations (753), and hence are given by

$$\left.\begin{aligned} z_{1kmm'}' &= m'z_{1km}'' \\ z_{2kmm'}' &= \frac{z_{2km}''}{m'} + \frac{1-m'^2}{4m'}z_{1km}'' \end{aligned}\right\}. \tag{790}$$

Substituting from (772a) this gives

$$\left.\begin{aligned} z_{1kmm'}' &= \cfrac{1}{\cfrac{1}{mm'z_{1k}} + \cfrac{1}{\cfrac{4mm'}{1-m^2}z_{2k}}} \\ z_{2kmm'}' &= \frac{z_{2k}}{mm'} + \cfrac{1}{\cfrac{1}{\cfrac{m(1-m'^2)}{m'(1-m^2)}z_{2k}} + \cfrac{1}{\cfrac{m(1-m'^2)}{4m'}z_{1k}}} \end{aligned}\right\}. \tag{790a}$$

These are also simple series and parallel combinations of portions of z_{1k} and z_{2k}. The manner in which this process is continued to determine the component reactances of succeeding derived types needs no further comment.

The propagation functions of the successive derivations are most

conveniently expressed in terms of the logarithmic forms (747) with the substitution (749). For the constant k this function is given by (747a). For the M-type of either sequence it is given by (752). This might have been written as

$$\gamma_{km} = \ln\left(\frac{y_{km} + 1}{y_{km} - 1}\right), \tag{752a}$$

where

$$y_{km} = \frac{y_k}{m}, \tag{791}$$

is obtained from (749) by replacing z_{1k} and z_{2k} by the component reactances for the M-type of either sequence.

For the MM'-type of either sequence we have by applying the same principles

$$\gamma_{kmm'} = \ln\left(\frac{y_{km} + m'}{y_{km} - m'}\right), \tag{792}$$

or

$$\gamma_{kmm'} = \ln\left(\frac{y_k + mm'}{y_k - mm'}\right). \tag{792a}$$

This may also be written as

$$\gamma_{kmm'} = \ln\left(\frac{y_{kmm'} + 1}{y_{kmm'} - 1}\right), \tag{792b}$$

whence

$$y_{kmm'} = \frac{y_{km}}{m'} = \frac{y_k}{mm'}. \tag{793}$$

Thus we have for either sequence

$$y_k : y_{km} : y_{kmm'} : \cdots = m : m' : m'' : \cdots. \tag{794}$$

These relations make it clear that for the propagation function of the derived type of any order we have

$$\gamma_{kg} = \ln\left(\frac{y_k + g}{y_k - g}\right), \tag{795}$$

where $$g = 1; m; mm'; mm'm''; \cdots, \tag{795a}$$

$g = 1$ representing the prototype or basic structure. In this sense we may generalize (749a) as

$$y_{kg} = \sqrt{1 - x_{kg}^{-2}}, \tag{796}$$

and obtain

$$x_{kg} = \frac{gx_k}{\sqrt{1 - (1 - g^2)x_k^2}}, \tag{797}$$

of which (755a) is a special case for $g = m$.

Since the derived type of any order has the same propagation function for either sequence we may extend the relation (773). For the MM'-type this reads

$$\frac{z_{1kmm'}{}'}{z_{2kmm'}{}'} = \frac{z_{1kmm'}{}''}{z_{2kmm'}{}''}. \tag{798}$$

For derived types of higher order similar relations hold. With (773a), (789), and (790) this gives

$$z_{1kmm'}{}' \, z_{2kmm'}{}'' = z_{1kmm'}{}'' \, z_{2kmm'}{}' = R^2. \tag{798a}$$

Similar relations are obtained for the higher orders.

To obtain the characteristic impedances of the MM'-type, we extend the relation (774) and have

$$\frac{W_{1km}}{W_{2kmm'}} = \frac{W_{1kmm'}}{W_{2km}} = 1 - x_{kmm'}{}^2. \tag{799}$$

Using (795a) and (797), and recalling (764) and (774a), this gives

$$\left.\begin{aligned} W_{1kmm'} &= \frac{W_{1k}\,\{1 - (1 - m^2)\,x_k^2\}}{1 - (1 - m^2m'^2)\,x_k^2} \\ &= \frac{R\sqrt{1 - x_k^2}\,\{1 - (1 - m^2)\,x_k^2\}}{1 - (1 - m^2m'^2)\,x_k^2} \end{aligned}\right\}, \tag{800}$$

and

$$\left.\begin{aligned} W_{2kmm'} &= \frac{W_{2k}\{1 - (1 - m^2m'^2)\,x_k^2\}}{1 - (1 - m^2)\,x_k^2} \\ &= \frac{R\{1 - (1 - m^2m'^2)\,x_k^2\}}{\sqrt{1 - x_k^2}\,\{1 - (1 - m^2)\,x_k^2\}} \end{aligned}\right\}. \tag{801}$$

The characteristic impedances for the derived types of higher order may be obtained in a similar manner. Each succeeding derivation introduces an additional factor of the form $\{1 - (1 - g^2)\,x_k^2\}$ alternately in the numerator and denominator of these expressions. The characteristic impedances of the various orders are either W_{1k} or W_{2k} times quotients of polynomials in x_k^2 whose coefficients are functions of m, m', etc. Over the range $-1 \leq x_k \leq 1$, such functions may be made to approximate a constant value with a tolerance and coverage which improve as the number of arbitrary coefficients increases, i.e., with the order of the derivation. Hence the degree of approximation may be made as close as desired by continuing to a sufficiently high order. The approximation problem is purely an algebraic one which we shall discuss further in what is to follow.

From what has been given it is seen that the relation (775) is extended to any order. For the MM'-type we have

$$W_{1kmm'} \cdot W_{2kmm'} = R^2, \tag{802}$$

Hence the derived type of either sequence possesses the same ability of approximation.

The attenuation function of the composite filter of either sequence is given by

$$\Gamma_1 = \ln\left|\frac{y_k + 1}{y_k - 1} \cdot \frac{y_k + m}{y_k - m} \cdot \cdot \cdot \cdot \frac{y_k + g}{y_k - g}\right|. \qquad (803)$$

In particular, for the sequence up to the second order we have

$$\Gamma_1 = \gamma_{1k} + \gamma_{1km} + \gamma_{1kmm'} = \ln\left|\frac{y_k + 1}{y_k - 1} \cdot \frac{y_k + m}{y_k - m} \cdot \frac{y_k + n}{y_k - n}\right|, \qquad (803a)$$

where we have put

$$n = mm'. \qquad (804)$$

The points of infinite attenuation lie at $y_k = 1$, $y_k = m$, and $y_k = n$. These are in the inverse order of their nearness to the cut-off, and are due, respectively, to the constant-k and the first and the second derived sections. Between these points of infinite attenuation lie two minima. The locations of these are found by setting the derivative of the argument of the logarithm in (803a) equal to zero. This gives the following biquadratic in y_k

$$(1 + m + n)\, y_k^4 - \{m^2 + n^2 + m(1 + n^2) + n(1 + m^2)\}\, y_k^2 + mn(m + n + mn) = 0. \qquad (805)$$

From considerations involving the approximation obtainable from $W_{1kmm'}$ and $W_{2kmm'}$, Zobel[1] chooses the parameter values $m = 0.7230$ and $m' = 0.4134$. With these, equation (805) becomes

$$2.022\, y_k^4 - 1.860\, y_k^2 + 0.2675 = 0,$$

from which

$$y_k = \pm\, 0.4220;\; y_k = \pm\, 0.8620,$$

corresponding to

$$x_k = \pm\, 1.103;\; x_k = \pm\, 1.972$$

respectively. Substituting back into (803a) we get for the minimum nearest the cut-off

$$\begin{aligned}\Gamma_{1\min} &= \ln\left(\frac{1.422}{0.578} \cdot \frac{1.145}{0.301} \cdot \frac{0.7209}{0.1231}\right) \\ &= \ln\,(2.46 \cdot 3.805 \cdot 5.855) \\ &= \ln 54.8 = 4.006 \text{ napiers} = 34.78 \text{ db.},\end{aligned}$$

[1] *Loc. cit.*, p. 313.

and for the second minimum

$$\Gamma_{1min} = \ln\left(\frac{1.862}{0.138} \cdot \frac{1.585}{0.139} \cdot \frac{1.161}{0.563}\right)$$
$$= \ln(13.48 \cdot 11.40 \cdot 2.062)$$
$$= \ln 317.0 = 5.76 \text{ napiers} = 50.02 \text{ db.}$$

The factors in the arguments of these logarithms are the contributions from the constant-k, the M, and the MM' sections, respectively. The first minimum is considerably smaller than the second but still high enough for many practical cases. A choice of a slightly larger m' will raise this considerably but will also affect the form of $W_{1kmm'}$ and $W_{2kmm'}$

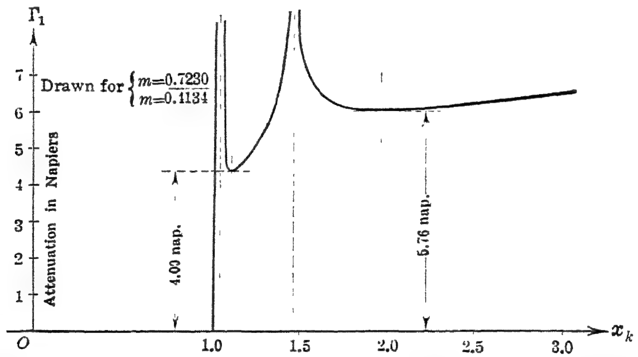


FIG. 128.—Attenuation of the composite filter consisting of a constant-k section cascaded with M- and MM'-type sections, as a function of x_k. Comparison with Fig. 127 shows the effect of adding a second derived section of the MM'-type.

as we shall show presently. Thus the choice of m-values still involves a compromise. This is an inherent feature of the design procedure upon which this type of filter is based.

The points of infinite attenuation in terms of x_k fall at

$$\left.\begin{aligned} x_k &= \pm\frac{1}{\sqrt{1-n^2}} = \pm 1.0475 \\ x_k &= \pm\frac{1}{\sqrt{1-m^2}} = \pm 1.454 \\ x_k &= \pm\infty \end{aligned}\right\}.$$

The situation in this numerical case is summarized in Fig. 128. Comparing this with Fig. 127 we see that the cut-off is very much sharper and the lowest minimum about the same.

Returning now to the characteristic impedance functions $W_{1kmm'}$ and $W_{2kmm'}$ given by (800) and (801), our problem is to determine how their

behavior in the transmission region depends upon the parameters m and m'. For certain ranges of values of these parameters the functions show maxima and minima within the region $-1 \leq x_k \leq 1$. Since both are even functions of x_k, only the region $0 \leq x_k \leq 1$ need be considered. Within this region $W_{1kmm'}$ has a maximum equal to R at $x_k = 0$, and may show a minimum and another maximum beyond this point. If R is chosen as the nominal value, i.e., that value which is to equal the terminal resistance, then such combinations of m and m' should be chosen which will cause the minimum and maximum beyond $x_k = 0$ to be equal deviations from the value R. It is then of interest to know how the maximum deviation from this mean value depends upon m and m', and what the corresponding coverage is. This may be done by, first, setting the derivative of the function equal to zero to find the locations of the maxima and minima and the conditions for which these lie within the prescribed interval, second, evaluating these maxima and minima and demanding that they shall be equal deviations from the mean value R, and finally determining the tolerance and coverage for a series of parameter values satisfying these conditions.[1] The results of such a process carried out with respect to $W_{1kmm'}$ are also applicable to $W_{2kmm'}$ with sufficient accuracy, and are summarized by Figs. 129 and 130. These are to be compared with Fig. 124 which applies to W_{1km} and W_{2km} on the same basis. It is apparent that considerably smaller tolerances and wider limits are obtained by the second derivation. Zobel's values of $m = 0.7230$ and $m' = 0.4134$ do not quite correspond to equal deviations from the mean value R but give rise to a tolerance of about 2 per cent over 96 per cent of the band. According to Figs. 129 and 130, the values $m = 0.707$ and $m' = 0.404$ correspond to a tolerance of 2 per cent over 96.5 per cent of the band. Fig. 131 illustrates $W_{1kmm'}$ for these values. For the plotting of such

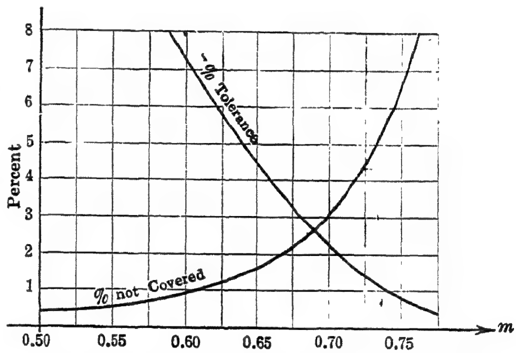


FIG. 129.—Tolerance and coverage for the characteristic impedance $W_{1kmm'}$ or $W_{2kmm'}$ in terms of the parameter m with the value R as a norm.

[1] For the details of such a process see "A Recent Contribution to the Design of Electric Filter Networks," Jour. of Math. and Phys., Vol. XI, No. 2, 1932, pp. 191, e.s.

curves it is helpful to know that the minimum and maximum beyond $x_k = 0$ occur for

$$x_k^2 = \frac{3a - a'}{2aa'}\left\{1 \pm \sqrt{1 - \frac{4aa'\,(1 + 2a - 2a')}{(3a - a')^2}}\right\}, \tag{806}$$

where

$$\left.\begin{aligned} a &= 1 - m^2 \\ a' &= 1 - m^2 m'^2 \end{aligned}\right\}. \tag{807}$$

With $m = 0.707$ and $m' = 0.404$ this gives

$$x_k = 0.643;\ 0.922.$$

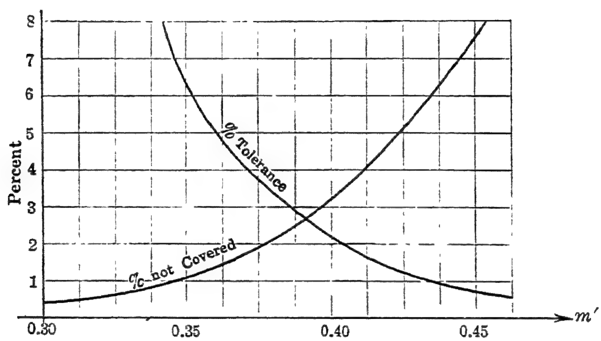


FIG. 130.—Tolerance and coverage for the characteristic impedance $W_{1kmm'}$ or $W_{2kmm'}$ in terms of the parameter m' with the value R as a norm.

Just as in the case of W_{1km} and W_{2km}, this method of terminating the MM'-type does not utilize the functions $W_{1kmm'}$ and $W_{2kmm'}$ to the best advantage. Somewhat smaller tolerances and wider limits may be

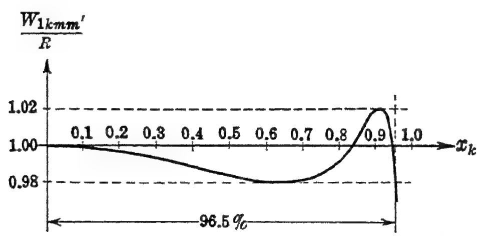


FIG. 131.—The impedance $W_{1kmm'}$ as a function of the frequency function x_k for the parameter values $m = 0.707$ and $m' = 0.404$. The value R is chosen as the norm.

obtained by fixing the parameters m and m' in such a way that the maximum of $W_{1kmm'}$ at $x_k = 0$ equals that beyond this point, and then terminating in a resistance equal to the mean value of this function.

For $W_{1kmm'}$ this value is somewhat less than R; for $W_{2kmm'}$ it is somewhat larger. Cauer[1] has stated how the parameters may be determined for given values of coverage and tolerance based upon the geometric mean values of these functions. Basing the tolerance upon the geometric instead of the arithmetic mean is really more proper because then the maximum positive and negative deviations from the mean correspond to the same magnitude for the reflection coefficient. This point is not very important, of course, since, for the small deviations involved here, there is little difference between a geometric and an arithmetic basis.

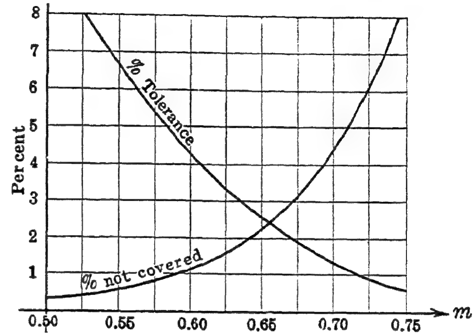


FIG. 132.—Plots corresponding to those of Fig. 129 but with the mean values (808), for $W_{1kmm'}$, or (809), for $W_{2kmm'}$ chosen as the norm.

The determination of m and m' on Cauer's basis for the above functions is summarized in Figs. 132 and 133. We see that the improvement is not very great as com-

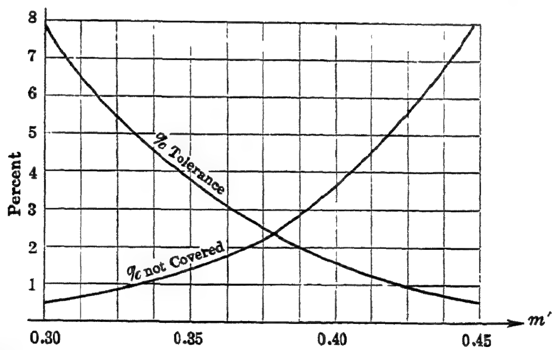


FIG. 133.—Plots corresponding to those of Fig. 130 but for the mean values referred to under Fig. 132 as a norm.

pared with the previous basis. For the values $m = 0.6710$ and $m' = 0.3886$, the tolerance is 2 per cent and the coverage 97 per cent, which is a little better than on the other basis. $W_{1kmm'}$ is illustrated for this

[1] W. Cauer, "Siebschaltungen," published by V.D.I., Verlag G.m.b.H., Berlin, 1931, pp. 6–8 and tables VI and VIII. The derivation of these results will be discussed in the following chapter.

case in Fig. 134. When the filter terminates with this characteristic impedance, the terminal resistance should be made equal to

$$R_t = \frac{R}{1 + \epsilon}. \tag{808}$$

When the filter terminates with $W_{2kmm'}$ we should have

$$R_t = R(1 + \epsilon). \tag{809}$$

Although the higher derived types offer a further improvement in both the attenuation and characteristic impedance functions, these are hardly worth treating in detail here since practical cases which call for such close limits seldom if ever occur. It must also be remembered in this connection that the full utilization of the functions presented by the MM' and higher order types requires that the network elements be held to very close limits. Practical difficulties encountered here tend to prohibit the use of types above MM'. In fact, the realization of the ultimate theoretical limits presents a problem in itself.

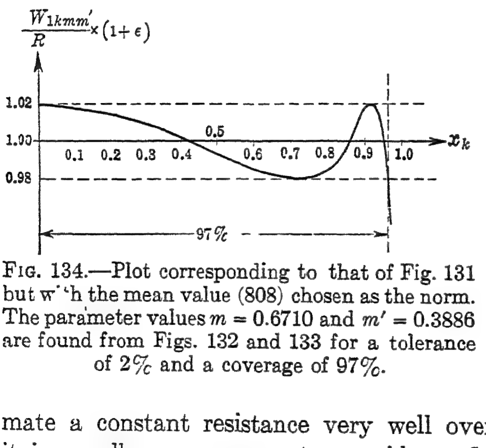


FIG. 134.—Plot corresponding to that of Fig. 131 but with the mean value (808) chosen as the norm. The parameter values $m = 0.6710$ and $m' = 0.3886$ are found from Figs. 132 and 133 for a tolerance of 2% and a coverage of 97%.

10. Reflection effects in the composite filter. Since the characteristic impedances W_{1kg} or W_{2kg} presented at the ends of a composite filter approximate a constant resistance very well over the transmission ranges, it is usually unnecessary to consider reflection effects within these. Where the filter attenuates, however, its characteristic impedance is imaginary so that reflection and interaction losses become more important. This situation was discussed in connection with the uniform ladder structure in section 5 of Chapter VIII. If we consider the composite filter to be working out of and into the nominal resistance R, then the logarithmic voltage ratio (658) becomes

$$\ln\left|\frac{E_g}{E_n}\right| = \ln 2 + \Gamma_1 - \ln|1 - r^2| + \ln|1 - r^2 e^{-2\Gamma_1}|, \tag{810}$$

where Γ_1 is the sum of the attenuation functions for the component sections, i.e.,

$$\Gamma_1 = \gamma_{1k} + \gamma_{1km} + \cdots + \gamma_{1kg}, \tag{811}$$

and r is the common reflection coefficient

$$r = \frac{R - W_{ikg}}{R + W_{ikg}}, \tag{812}$$

in which the subscript i is 1 or 2 according to whether the termination is mid-series or mid-shunt, respectively.

Letting

$$u = \frac{W_{ikg}}{jR}, \tag{813}$$

this reflection coefficient becomes

$$r = \frac{1 - ju}{1 + ju}, \tag{812a}$$

and we find

$$|1 - r^2| = \rho = \frac{4|u|}{1 + u^2}. \tag{814}$$

Applying the same process as was used in the derivation of (669), we find here for the approximate interaction loss

$$\ln|1 - r^2 e^{-2\Gamma_1}| = - e^{-2\Gamma_1}\left(1 - \frac{\rho^2}{2}\right). \tag{815}$$

This is usually negligible except in the immediate vicinity of the cut-off. At the cut-off the last two terms in (810) become infinite and opposite and their combined effect indeterminate. If for the moment we let $h = e^{-\Gamma_1}$, then the last two terms in question are found equivalent to

$$\frac{1}{2}\ln\left\{\left(\frac{1 + h^2}{2}\right)^2 + \frac{(1 - u^2)^2(1 - h^2)^2}{16u^2}\right\}. \tag{816}$$

Using (803) for Γ_1, and expressing u in terms of y_k also, the value at any cut-off is given by taking the limit $y_k \to 0$. For the filter of either sequence involving the single m-derivation this gives

$$\tfrac{1}{2}\ln\{1 + m^2(1 + m)^2\}, \tag{816a}$$

while for either sequence up to the second derivation we have

$$\frac{1}{2}\ln\left\{1 + \left(\frac{m'}{m}\right)^2(1 + m' + mm')^2\right\}. \tag{816b}$$

Where the attenuation is sufficient to make h^2 negligible in comparison with unity, (816) reduces to the reflection loss $-\ln\rho$ as it obviously should.

For purposes of computing the reflection loss it is convenient to plot ρ versus u. Such a plot may be used for filters of all classes and orders. This is shown in Fig. 135. We see that for values of u between $2 \pm \sqrt{3}$ the reflection loss is negative but that its maximum negative value is restricted to $\ln 2 = 0.693$ napier $= 6.02$ decibels, the latter occurring at the point where $u = 1$ which, according to (813), corresponds to those frequencies at which the magnitude of the characteristic impedance equals the nominal resistance R. Wherever this impedance is either zero or infinite, u is zero or infinite and hence (with the exception of the boundaries) the loss due to reflection effects is infinite. According to a generalization of (800) and (801) these points correspond to frequencies where

$$x_k = \frac{1}{\sqrt{1 - g^2}};\ g = 1,\ m,\ mm',\ \cdots \text{etc.}$$

By means of (749a) and (803) these are seen to be those frequencies at which Γ_1 is infinite. Thus, although the reflection effects introduce

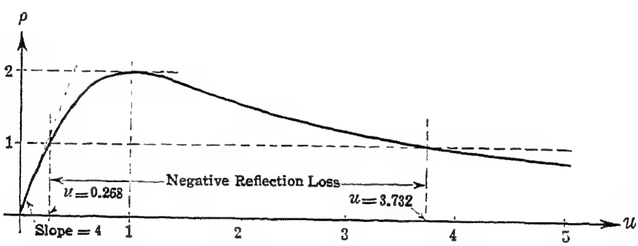


FIG. 135.—The reflection factor (814) in terms of the function defined by eq. (813).

points of infinite attenuation, these coincide with those already provided by the attenuation function and hence no additional infinite maxima are produced. In the intervening regions where Γ_1 subsides to its minimum values, the reflection loss becomes negative and tends to lower these minima. The reflection effects, therefore, do not cooperate in providing the desired net attenuation. They aid the attenuation function where it needs no aid, and oppose it where support would be desirable.

For purposes of calculation it is useful to note that the relation (814) for ρ is unchanged if u is replaced by $1/u$, which is equivalent to saying that Fig. 135 exhibits geometric symmetry about the point $u = 1$. As a result of this property the reflection loss is the same for mid-series and mid-shunt terminations of the same order. This becomes clear if we recall that W_{1k_g} and W_{2k_g} are inverse with respect to R^2, and hence lead to inverse values for u according to (813).

Fig. 136 shows a plot of the reflection and interaction loss versus x_k for the composite filter of either sequence up to the single m-derivation,

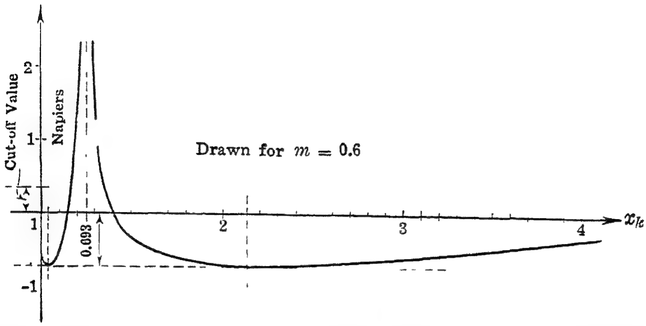


FIG. 136.—Reflection and interaction losses versus the frequency function x_k for the composite filter of Fig. 126 (or the corresponding one in the shunt sequence) for terminal resistances equal to the nominal value R.

and Fig. 136a shows how these reflection effects influence the net attenuation. The actual minimum is about one-half a napier less than that of Γ_1. The sharpness of the cut-off is not affected by any appre-

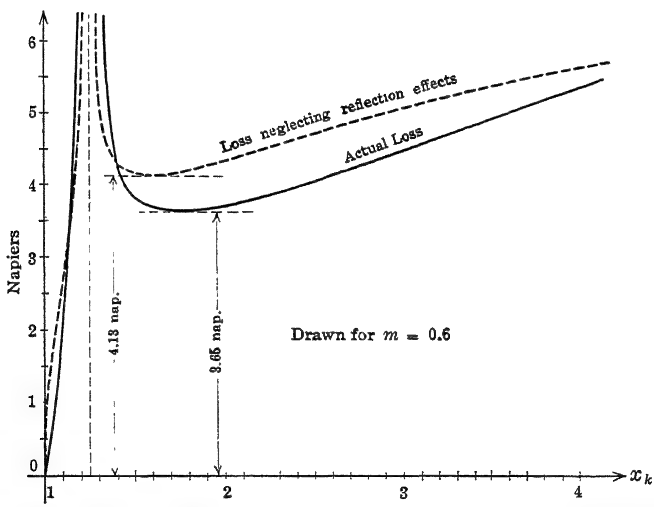


FIG. 136a.—Effect of reflection and interaction losses upon the attenuation of the composite filter consisting of a constant-k and an M-type section. Resultant curve is the sum of curves in Figs. 127 and 136.

ciable amount. Fig. 136 shows, incidentally, that the combined reflection and interaction loss at the cut-off, as evaluated by (816a), is hardly noticeable since the curve drops so rapidly at this point.

11. Fractional terminations. It frequently occurs in practice that more than one filter is to work out of the same source. This happens, for example, when several carrier channels coming over the same facility are to be separated. Such a situation may be met by placing the various filters either in series or in parallel on their input sides and connecting the resulting combination to the common source. Here a difficulty arises, however. Unless special precautions are taken, the filters will interfere with each other's behavior so as to produce undesirable results. The reason for this lies in the fact that the individual characteristic impedances are real and approximately constant only over the transmission regions of their respective filters. In the attenuation regions they are imaginary[1] and behave like physical reactances; i.e., they possess alternate poles and zeros. Since the transmission region of one filter is the attenuation region of another, this means that, if the networks are in parallel, the zeros of the characteristic impedance of one effectively short-circuit the input of the others, and if the networks are in series, the poles of one effectively open-circuit the input to all. At intermediate frequencies the individual behaviors are also considerably interfered with. The solution lies in devising some method whereby the combined susceptance of all parallel networks, or the combined reactance of all series networks, is annulled over their total transmission range. This process we wish to discuss now.

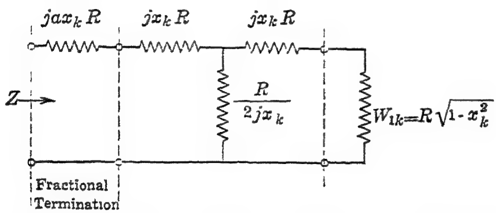


FIG. 137.—The properly terminated constant-k T-section with a fractional series branch on the input side.

As an introduction to a general solution of our present problem we shall study the characteristic impedance of a uniform ladder structure of the constant-k type with a fractional series termination. By this we mean the impedance Z looking into the structure of Fig. 137. The particular expressions used here for the component reactances z_{1k} and z_{2k} follow from (682) and (684), that is

$$\left.\begin{aligned} z_{1k} &= 2jx_kR \\ z_{2k} &= \frac{R}{2jx_k} \end{aligned}\right\}. \tag{817}$$

The factor a in Fig. 137 is a numeric which may be positive or negative according to whether we wish to consider the structure as beginning

[1] Dissipation is still neglected in this discussion.

with *more* or *less* than half of z_{1k}. For $a = 0$ we get the familiar case $Z = W_{1k}$. In general we have

$$Z = R\{\sqrt{1 - x_k^2} + jax_k\}\cdot \tag{818}$$

For our further considerations it is more useful to study the corresponding admittance function

$$Y = \frac{1}{Z} = G + jB, \tag{819}$$

and in particular its real and imaginary parts G and B. From (818) ve get

$$R(G + jB) = \frac{1}{\sqrt{1 - x_k^2} + jax_k}\cdot \tag{820}$$

Since the factor $\sqrt{1 - x_k^2}$ may be either real or imaginary according to whether we consider the transmission or attenuation region, respectively, we must evidently study the function in these ranges separately. Here we are further reminded of the fact brought out earlier that the radical takes the minus sign for $-\infty < x_k \leq -1$, and the plus sign for $+1 \leq x_k < \infty$.

Thus we find *for the range* $-\infty < x_k \leq -1$

$$RG = 0;\ RB = \frac{1}{\sqrt{x_k^2 - 1} - ax_k}, \tag{821}$$

for the range $-1 \leq x_k \leq 1$

$$\left.\begin{aligned} RG &= \frac{\sqrt{1 - x_k^2}}{1 - (1 - a^2)\,x_k^2} \\ RB &= \frac{-\,ax_k}{1 - (1 - a^2)\,x_k^2} \end{aligned}\right\}, \tag{821a}$$

and for the range $1 \leq x_k < \infty$

$$RG = 0;\ RB = \frac{-1}{\sqrt{x_k^2 - 1} + ax_k}\cdot \tag{821b}$$

This is summarized by Fig. 138 which is drawn for $a = 0.6$. Several points are of interest here. First, we note that the conductance RG has the same appearance in the transmission range as W_{1km}/R, as is evident from a comparison of the first relation (821*a*) and (774*a*). The only difference is that a replaces m. Second, we see that the susceptance RB is an odd function of x_k and that it has a negative slope in the transmission region. This means that it may be partly or wholly canceled within this range by means of a physical susceptance with positive

slope. This may be accomplished by placing across the input terminals of the network of Fig. 137 a two-terminal network with the reactance

$$\frac{-1}{jB}.$$

Substituting from (821a), this so-called **susceptance-annulling network** should have a reactance given by

$$\frac{R\{1-(1-a^2)\,x_k^2\}}{jax_k} = \frac{R}{jax_k} + \frac{(1-a^2)}{a}jx_kR$$
$$= 2\left\{\frac{z_{2k}}{a} + \frac{1-a^2}{4a}z_{1k}\right\}. \qquad (822)$$

Recalling (753), we see that this is the same as $2z_{2km}'$ except that a replaces m. Incidentally the fractional series reactance, which in Fig.

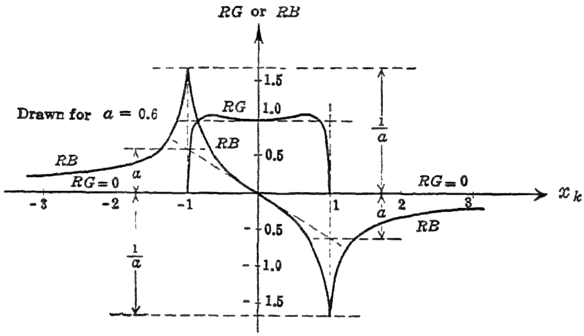


FIG. 138.—The conductance and susceptance functions corresponding to the impedance Z (divided by R) of Fig. 137 plotted versus the frequency function x_k.

137 is denoted by jax_kR, is recognized as equivalent to $\frac{1}{2}az_{1k}$ and hence the same as $\frac{1}{2}z_{1km}'$. Thus the fractional termination with its susceptance-annulling network is identical with a half-section of the mid-series derived filter. The present procedure so far is merely an alternative way of deriving this result.

Our immediate problem, however, leads us to study from another angle the situation revealed by Fig. 137. From the fact that the susceptance RB has a negative slope in the transmission range and a positive one in the attenuation range, we recognize that if two filters with fractional series terminations are placed in parallel on their input sides they may become each other's susceptance-annulling networks if their transmission and attenuation ranges mutually replace each other. Filters which are related in this way are said to be **complementary.** Low-pass and high-pass, or band-pass and band-elimination filters with

identical cut-off frequencies are complementary. If for one such filter x_k equals say x_{k1} and for the other $x_k = x_{k2}$, then the general definition for a pair of complementary filters may be expressed as

$$x_{k1} \cdot x_{k2} = -1. \tag{823}$$

The relations (725) and (742) illustrate this for the **L.P.-H.P.** and **B.P.-B.E.** classes, respectively.

Although the paralleling of complementary filters cannot result in complete mutual susceptance-annulling over their combined transmission ranges, yet a fairly good result may be obtained by a proper choice of the parameter a. Let us consider, for example, the **L.P.** and **H.P.** classes. For these x_k equals ω/ω_c and $-\omega_c/\omega$, respectively. Thus the application of (821a) and (821b) to the low-pass gives

$$\left.\begin{aligned} RB &= \frac{-a\left(\dfrac{\omega}{\omega_c}\right)}{1-(1-a^2)\left(\dfrac{\omega}{\omega_c}\right)^2};\ 0 \leq \frac{\omega}{\omega_c} \leq 1 \\ RB &= \frac{-1}{\sqrt{\left(\dfrac{\omega}{\omega_c}\right)^2 - 1} + a\left(\dfrac{\omega}{\omega_c}\right)};\ 1 \leq \frac{\omega}{\omega_c} < \infty \end{aligned}\right\}, \tag{824}$$

and the application of (821) and (821a) to the high-pass gives

$$\left.\begin{aligned} RB &= \frac{1}{\sqrt{\left(\dfrac{\omega_c}{\omega}\right)^2 - 1} + a\left(\dfrac{\omega_c}{\omega}\right)};\ 0 \leq \frac{\omega}{\omega_c} \leq 1 \\ RB &= \frac{a\left(\dfrac{\omega_c}{\omega}\right)}{1-(1-a^2)\left(\dfrac{\omega_c}{\omega}\right)^2};\ 1 \leq \frac{\omega}{\omega_c} < \infty \end{aligned}\right\}. \tag{825}$$

These are plotted in Fig. 139 for $a = 0.6$. The dotted curve shows the residual susceptance which is quite small except in the vicinity of the cut-off. This curve has negative geometric symmetry about the point $x_k = 1$. By this we mean, for example, that the ordinate for $x_k = \frac{1}{2}$ is the negative of that for $x_k = 2$. A somewhat smaller value of a will reduce the maximum deviations in the vicinity of the cut-off but will introduce other extrema, one nearer the origin and another at a geometrically located point. The resulting behavior of these paralleled filters with $a = 0.6$ is quite satisfactory. Fig. 140 shows the combined network with a mid-series derived half-section on the output end of each filter. The half-sections are determined for $m = 0.6$.

The fact that these filters act as each other's susceptance-annulling networks effects an economy in the total number of coils and condensers involved as compared with the case where each filter operates separately.

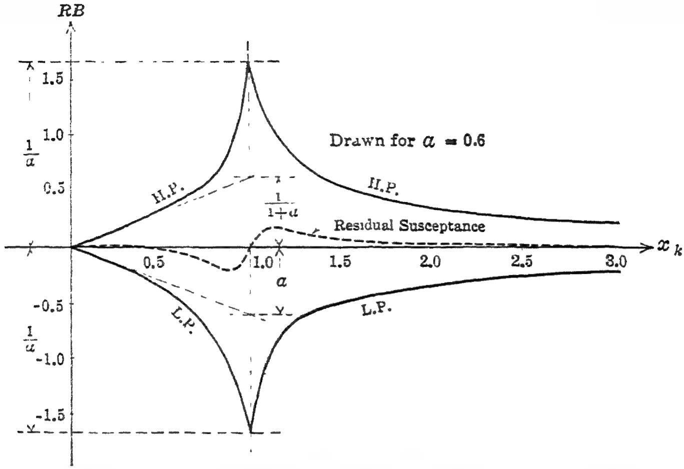


FIG. 139.—Susceptance characteristics for a pair of complementary filters with fractional series terminations according to Figs. 137 and 138.

In the parallel combination shown in Fig. 140, each filter shows very nearly the same behavior regarding terminal conditions as would separate constant-k filters with m-derived half-sections symmetrically placed at either end. The propagation functions are not the same, however, since the paralleled filter does not replace the shunt reactance z_{2km}' for its mate in the latter's attenuation range.

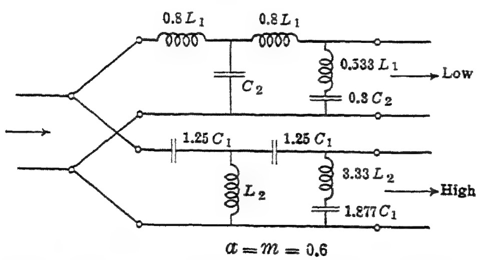


FIG. 140.—Low-pass and high-pass constant-k filters paralleled by means of fractional series terminations on their input sides. The output sides are individually supplied with M-type half-sections to simulate composite filter behavior.

This method of operating complementary filters in parallel is due to Zobel, who calls the fractional series reactances "x-terminations."[1] Fig. 139 applies to complementary filters of all classes, and hence the above process may be extended to other pairs of filters with the same resulting behavior.

[1] See U. S. Patents Nos. 1,557,229 and 1,557,230.

When the filters are potentially complementary but do not have identical cut-off frequencies, then a somewhat different procedure becomes necessary. This we shall discuss for the **L.P.** and **H.P.** case. Let the cut-off frequencies for these filters be ω_{cl} and ω_{ch}, respectively, with $\omega_{cl} < \omega_{ch}$. The susceptance characteristics RB for this combination are shown in Fig. 141. It is evident that they do not cancel each other very well. Hence it becomes necessary to place a susceptance in parallel with the input sides of the filters which is equal and opposite to the difference of the susceptances over the pass bands. It is, of course, only necessary to annul the resulting susceptance over the pass bands.

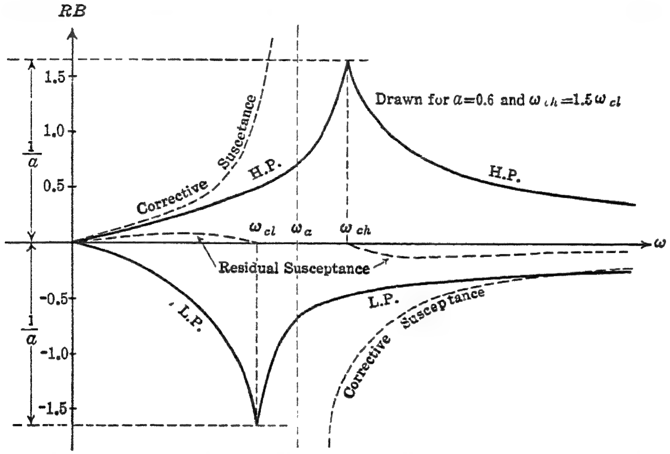


FIG. 141.—Susceptance characteristics and corrective susceptance for a pair of potentially complementary filters (for example, low-pass and high-pass) with fractional series terminations but different cut-off frequencies.

Within the interval $\omega_{cl} < \omega < \omega_{ch}$ the net susceptance may be anything at all.

The simplest susceptance-annulling network in this case is a resonant component. This has the susceptance characteristic shown by the dotted curve of Fig. 141. Denoting the inductance and capacitance in this resonant component by L_a and C_a, and letting

$$\omega_a = \frac{1}{\sqrt{L_a C_a}} \tag{826}$$

be the resonant frequency, the susceptance of the annulling network when multiplied by R becomes

$$RB_a = \frac{R\omega}{L_a(\omega_a^2 - \omega^2)}. \tag{827}$$

If the positive ordinate of this function at ω_{cl} is to equal the negative ordinate at ω_{ch}, then

$$\omega_a = \sqrt{\omega_{cl}\,\omega_{ch}}. \tag{826a}$$

This will give rise to the same degree of approximation in both pass bands.

The parameters in the annulling network are then completely determined by one more condition. Suppose we demand that the correction shall be complete at the cut-off frequencies ω_{cl} and ω_{ch}. Writing this condition at ω_{cl} with the help of the first relation (825) and (827) we have

$$\frac{1}{\sqrt{\left(\frac{\omega_{ch}}{\omega_{cl}}\right)^2 - 1} + a\left(\frac{\omega_{ch}}{\omega_{cl}}\right)} + \frac{R\omega_{cl}}{L_a(\omega_a^2 - \omega_{cl}^2)} = \frac{1}{a}, \tag{828}$$

from which

$$L_a = \frac{aR}{(\omega_{ch} - \omega_{cl})} \cdot \frac{\sqrt{\omega_{ch}^2 - \omega_{cl}^2} + a\omega_{ch}}{\sqrt{\omega_{ch}^2 - \omega_{cl}^2} + a(\omega_{ch} - \omega_{cl})}. \tag{828a}$$

The corrective characteristic in Fig. 141 is drawn for this value. The residual susceptance is quite small. It could be made to average a smaller value by determining L_a so as to make the correction complete at a frequency somewhat below ω_{cl} and above ω_{ch}.

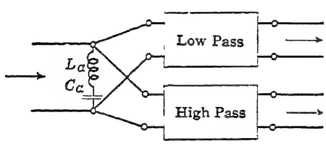


FIG. 142.—Susceptance annulling network in parallel with the filters whose characteristics are shown in Fig. 141.

The susceptance-annulling network is bridged across the common input to the two filters as shown in Fig. 142. The same process may be applied to filters of other classes. For a pair of band-pass and band-elimination filters having the same geometric mean frequency but a wider suppression than a pass band, the simplest annulling network consists of two resonant components in parallel. The susceptance characteristic of this network has a zero at the common geometric mean frequency. The two poles lie between the cut-off frequencies on either side. The parameters are most readily found by a cut-and-try process with the aid of Foster's reactance theorem.

A case of considerable practical importance is that involving the paralleling of several band-pass filters having adjacent pass bands. The susceptance characteristics for these are determined with the help of the relations (821) to (821b) and equation (730a), page 319, which

gives x_k for the band-pass filter. The exact shape of these curves depends upon the ratio of band width to the geometric mean between the cut-off frequencies for the individual filters. If this ratio is small, and the filters have identical band widths, then the individual susceptance characteristics become very nearly identical except for a horizontal displacement equal to the band width.

This may be seen if we expand the expression (730a) for x_k in terms of the ratio $(\omega - \omega_0)/\omega_0$ so as to get

$$x_k = \frac{2(\omega - \omega_0)}{w}\left\{1 - \frac{\omega - \omega_0}{2\omega_0} + \frac{(\omega - \omega_0)^2}{2\omega_0^2} - \cdots\right\}. \tag{829}$$

The region of interest is usually confined to the vicinity of $\omega = \omega_0$. If the largest value of $(\omega - \omega_0)/2\omega_0$ is a decimal error which may still be

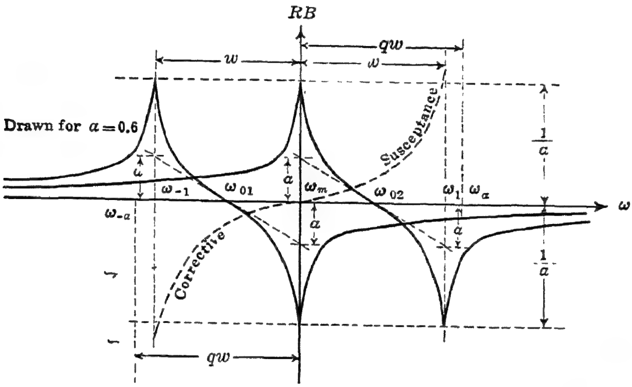


FIG. 143.—Susceptance characteristics and corrective susceptance for a pair of band-pass filters whose pass-bands are adjacent

tolerated, we can use the approximate expression for the band-pass filter

$$x_k \cong \frac{2(\omega - \omega_0)}{w}. \tag{829a}$$

Then x_k becomes a linear function of frequency, and the susceptance characteristics have the shape given in Fig. 138.

Fig. 143, which is drawn for this condition, illustrates the situation for two filters with identical band widths. The dotted curve is that susceptance characteristic which will completely annul the net susceptance of the filters over their combined pass bands.The figure is drawn for fractional series terminations with $a = 0.6$. To a first approximation the susceptance-annulling network is given by an anti-resonant com-

ponent, i.e., a coil and condenser in parallel, for which the anti-resonant frequency is that dividing the two pass bands.

A very much better result is obtained by means of two resonant components in parallel, the susceptance of which has a zero at the boundary between the pass bands and poles just beyond the boundaries at either end. This susceptance is given by

$$Ry_c = \frac{-HR\omega(\omega^2 - \omega_m{}^2)}{(\omega^2 - \omega_{-a}{}^2)(\omega^2 - \omega_a{}^2)}, \tag{830}$$

where the notation is as indicated in Fig. 143. The determination of the corresponding network according to either Foster's or Cauer's methods requires the knowledge of H, ω_a, and ω_{-a}. This may be done by a cut-and-try process in the following manner:

Where the approximation (829a) is allowed, we may write

$$\left.\begin{aligned} \omega^2 - \omega_m{}^2 &\cong 2\omega_m(\omega - \omega_m) \\ (\omega^2 - \omega_{-a}{}^2)(\omega^2 - \omega_a{}^2) &\cong 4\omega_m{}^2(\omega - \omega_{-a})(\omega - \omega_a) \end{aligned}\right\}, \tag{831}$$

and fix ω_{-a} and ω_a at points equidistant from ω_m so that

$$\left.\begin{aligned} \omega_{-a} &= \omega_m - qw \\ \omega_a &= \omega_m + qw \end{aligned}\right\}, \tag{832}$$

where q is a numeric larger than unity. The susceptance (830) then becomes

$$Ry_c \cong -\frac{HR\omega}{2\omega_m} \cdot \frac{(\omega - \omega_m)}{(\omega - \omega_m + qw)(\omega - \omega_m - qw)}, \tag{830a}$$

which shows negative symmetry about ω_m.

The quantities H and q may be determined by fixing the value of the function (830a) at two arbitrary points. On account of the symmetry these are advantageously chosen on the same side of ω_m. Suppose we choose them at ω_1 and midway between ω_{02} and ω_1. Demanding complete agreement between (830a) and the required corrective susceptance at these points will result in fairly good agreement over the combined pass bands. Determining the values of the corrective susceptance at these points by means of the second relation (821a) and (821b) with ω_{02} and ω_{01}, respectively, for ω_0 in (829a), we have

$$Ry_{c(\omega_1)} \cong \frac{HR}{2w(q^2 - 1)} = 1.882$$

$$Ry_{c(\omega_m + \frac{3}{4}w)} \cong \frac{3HR}{8w(q^2 - \frac{9}{16})} = 0.621$$

from which we find

$$q = 1.16;\ HR = 1.3w. \tag{833}$$

These together with (832) determine the necessary parameters in (830) for the evaluation of the annulling network when w is given.

When the band width is not narrow compared with its distance from the origin, the corrective characteristic in Fig. 143 is not symmetrical, but the same general procedure is applicable for the determination of the annulling network. The above figures may then serve as a starting point in a more extensive cut-and-try process. Any desired degree of approximation may be attained by introducing additional poles and zeros in the susceptance function of the annulling network. These additional resonant and anti-resonant frequencies (which are chosen to lie outside the pass bands, of course) serve as design parameters by means of which the degree of approximation may be controlled.

The same general picture also applies to the case of more band-pass filters in parallel. The student may illustrate these for himself as practice problems.

Before concluding this discussion we wish to point out that substantially the same results may be obtained by means of fractional shunt terminations on mid-shunt sections. In that case the argument is carried out in terms of reactances instead of susceptances. The filters are connected in series on their input sides and a reactance-annulling network is placed in series with the combination. This network serves to correct the net reactance of all filters over their combined transmission ranges.

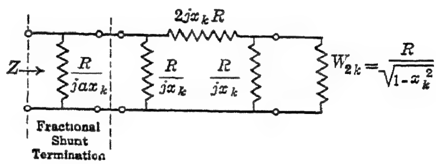


FIG. 144.—Properly terminated mid-shunt constant-k section with fractional shunt branch on the input side.

For a single mid-shunt section the fractional termination is illustrated by Fig. 144. The impedance looking into the left-hand terminals is

$$Z = \frac{R}{\sqrt{1 - x_k^2} + jax_k}. \tag{834}$$

Writing

$$\frac{Z}{R} = \frac{\mathcal{R} + j\mathcal{X}}{R} = \frac{1}{\sqrt{1 - x_k^2} + jax_k}, \tag{834a}$$

we see that this is identical to (820). Hence if in the above discussions we replace RG by $\mathcal{R}/R$ and RB by $\mathcal{X}/R$, the results apply here with the exception that the filters and the annulling network are placed in series. Hence the discussion of this case need be carried no farther. Fig. 145 shows a pair of complementary low- and high-pass filters connected on this basis with shunt-derived half-sections on their output ends. The

parameters a and m are taken equal to 0.6, so that the net behavior is practically the same as for the parallel combination of Fig. 140.

12. Impedance correction. The method discussed in the previous section involving the use of fractional series or shunt terminations forms the first step in a building-out process suggested by Bode[1] for the purpose of accomplishing substantially the same results as Zobel obtains by means of his m-derivations. We saw above that the first fractional termination plus its susceptance- or reactance-annulling network leads to the same network as a half-section of the series or shunt derived type. The continuation of Bode's method, although it may be made to yield similar end-results as are obtained with the addition of higher m-derived half-sections, takes a somewhat different form both physically as well as analytically. This we wish to discuss briefly.[2]

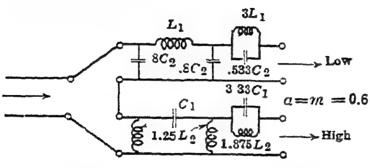


FIG. 145.—Low-pass and high-pass filters with fractional shunt branches on their input sides connected in series for band separation from a common source. This is the dual arrangement to that of Fig. 140.

Fig. 146 illustrates one way in which the corrective procedure may be laid out. R/y is the impedance to be corrected. This may be the mid-series characteristic impedance of a constant-k filter, in which case

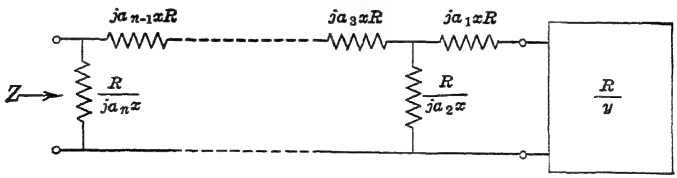


FIG. 146.—Terminating structure for Bode's method of impedance correction.

$x = x_k$. The terminating structure consists of alternate series and shunt arms. The series arms are all proportional to an impedance jxR, the proportionality factors being $a_1, a_3, \ldots a_{n-1}$, while the shunt arms are

[1] H. W. Bode, "A Method of Impedance Correction," *B.S.T.J.*, Vol. IX, pp. 794–835, October, 1930.

[2] An essential difference between the method of m-derivation and the method of impedance correction discussed in this section lies in the fact that the terminating network in the latter method does not match the filter on the image basis as do the m-derived structures. This makes the analysis of the structure less convenient and causes a reflection effect at the junction which may, however, be neutralized by corresponding variations in the image impedance at the other end of the filter.

proportional to R/jx, the reciprocal factors being $a_2, a_4, \ldots a_n$. The method of the preceding section utilizes only the first arm of this structure. There we saw that the addition of this fractional arm led to an improved or partially corrected net conductance. Here we shall see that, by continuing the building-out process, the net conductance may be further improved, the degree of improvement depending upon the extent to which the process is carried.

Just as in the case of the single fractional arm, however, this terminating structure serves to correct only the conductance part of the given admittance. The net susceptance must then be separately corrected by finally placing a special two-terminal network in parallel with the input terminals. The same network can be used, of course, to correct the resistance part of the given impedance, in which case the net reactance must be corrected by a final series network. The argument in general can be carried out in terms of resistance and reactance as well as conductance and susceptance. Specifically, however, when the last arm in the corrective structure is shunt, as in Fig. 146, it is more effective from the standpoint of the degree of correction obtainable to argue in terms of resistance and reactance, for, if the argument were then carried on the basis of conductance, the last shunt arm could have no corrective effect since it is itself a susceptance. Similarly, when the last arm is a series branch, the corrective procedure should be argued in terms of conductance. The first branch in the corrective structure should be series or shunt according to whether it attaches to a mid-series or mid-shunt terminated ladder network. In the following we shall discuss only the situation as illustrated in Fig. 146. Other modifications which may be required in specific cases follow the same general procedure.

The impedance Z looking into the left-hand terminals of the network of Fig. 146 is easily seen to be given by the following continued fraction:

$$Z = \frac{1 \,|}{ja_n x/R} + \frac{1 \,|}{|ja_{n-1} xR} + \cdots + \frac{1 \,|}{|ja_2 x/R} + \frac{1 \,|}{|ja_1 xR} + \frac{1}{|y/R}. \qquad (835)$$

Multiplying through by $1/R$ this is

$$\frac{Z}{R} = \frac{1 \,|}{ja_n x} + \frac{1 \,|}{|ja_{n-1} x} + \cdots + \frac{1 \,|}{|ja_2 x} + \frac{|}{|ja_1 x} + \frac{1}{|y}. \qquad (835a)$$

Here it is convenient to introduce the notation

$$\alpha_k = ja_k x, \qquad (836)$$

so that we have

$$\frac{Z}{R} = \frac{1|}{\alpha_n} + \frac{1 \,|}{|\alpha_{n-1}} + \cdots + \frac{1|}{|\alpha_2} + \frac{1|}{|\alpha_1} + \frac{1}{|y}. \qquad (835b)$$

A finite continued fraction of this sort is most conveniently treated in terms of what are known as *continuants*.[1] By means of these the continued fraction

$$d_1 + \frac{1|}{|d_2} + \frac{1|}{|d_3} + \cdots + \frac{1\,|}{|d_{n-1}} + \frac{1}{|d_n} \tag{837}$$

is written

$$\frac{K(d_1, d_2, \ldots d_n)}{K(d_2, d_3, \ldots d_n)}, \tag{837a}$$

where the continuants $K(\ldots)$ are defined by the recursion formula

$$K(d_1, d_2, \ldots d_n) = d_n\, K(d_1, d_2, \ldots d_{n-1}) + K(d_1, d_2, \ldots d_{n-2}), \tag{838}$$

with

$$K(d_1) = d_1\ ;\ K(0) = 1. \tag{838a}$$

The reader may very easily verify this by noting the manner in which the following series of simple cases lead to the recursion formula (838) and the initial values (838a)

$$d_1 = \frac{K(d_1)}{K(0)},$$

$$d_1 + \frac{1}{|d_2} = \frac{d_2 d_1 + 1}{d_2} = \frac{K(d_1, d_2)}{K(d_2)},$$

$$d_1 + \frac{1|}{|d_2} + \frac{1}{|d_3} = \frac{d_3(d_2 d_1 + 1) + d_1}{d_3 d_2 + 1}$$

$$= \frac{K(d_1, d_2, d_3)}{K(d_2, d_3)}, \text{ etc.}$$

Thus the impedance (835b) may be written

$$\frac{Z}{R} = \frac{K(\alpha_{n-1}, \ldots \alpha_1, y)}{K(\alpha_n, \ldots \alpha_1, y)}. \tag{839}$$

This may be put into an alternative form by utilizing the continuant property

$$K(d_1, d_2, \ldots d_n) = K(d_n, d_{n-1}, \ldots d_1), \tag{840}$$

which enables the recursion formula (838) to be written

$$K(d_1 \ldots d_n) = d_1 K(d_2 \ldots d_n) + K(d_3 \ldots d_n). \tag{838b}$$

Hence we can also write

$$\frac{Z}{R} = \frac{K(y, \alpha_1, \ldots \alpha_{n-1})}{K(y, \alpha_1, \ldots \alpha_n)}, \tag{839a}$$

[1] See Chrystal's "Text Book of Algebra," Adam and Charles Black, Edinburgh, 1889, Chapter XXXIV, pp. 463 e.s.; or Muir's "Theory of Determinants," Macmillan and Co., London, 1882, Chapter III, pp. 149–161.

and applying (838*b*) this is

$$\frac{Z}{R} = \frac{yK(\alpha_1 \ldots \alpha_{n-1}) + K(\alpha_2 \ldots \alpha_{n-1})}{yK(\alpha_1 \ldots \alpha_n) + K(\alpha_2 \ldots \alpha_n)}. \tag{839b}$$

Since a continuant whose argument contains an odd number of elements is given by terms which are products of odd numbers of elements only, while one whose argument contains an even number of elements is given by terms involving only even numbers of elements, the imaginary α's will cause the continuants in (839*b*) to be real or imaginary according to whether their arguments contain even or odd numbers of elements, respectively. For the structure of Fig. 146, n is an *even* integer. Hence $K(\alpha_1 \ldots \alpha_{n-1})$ and $K(\alpha_2 \ldots \alpha_n)$ are imaginary while $K(\alpha_2 \ldots \alpha_{n-1})$ and $K(\alpha_1 \ldots \alpha_n)$ are real.

In order to conserve space let us omit writing the arguments for these continuants and simply attach the subscripts of the first and last α's to K, for example

$$K(\alpha_1 \ldots \alpha_{n-1}) = K_{1,n-1}.$$

Then if we assume that the admittance y may be complex in general so that

$$y = y_1 + jy_2,$$

the impedance (839*b*) becomes

$$\frac{Z}{R} = \frac{(K_{2,n-1} + jy_2K_{1,n-1}) + y_1K_{1,n-1}}{y_1K_{1,n} + (K_{2,n} + jy_2K_{1,n})}, \tag{841}$$

where the first terms in the numerator and denominator are real and the second are imaginary.

Rationalizing by multiplying numerator and denominator by the conjugate denominator, we find

$$\frac{\mathcal{R} + j\mathcal{X}}{R} = \frac{\{(K_{2,n-1} + jy_2K_{1,n-1}) + y_1K_{1,n-1}\}\{y_1K_{1,n} - (K_{2,n} + jy_2K_{1,n})\}}{y_1^2K_{1,n}^2 - (K_{2,n} + jy_2K_{1,n})^2}. \tag{841a}$$

Utilizing the continuant property

$$K_{1,n}K_{2,n-1} - K_{1,n-1}K_{2,n} = (-1)^n, \tag{842}$$

we get for the real part

$$\frac{\mathcal{R}}{R} = \frac{y_1}{y_1^2K_{1,n}^2 - (K_{2,n} + jy_2K_{1,n})^2}, \tag{843}$$

while for the imaginary part we have

$$\frac{j\mathcal{X}}{R} = \frac{(y_1^2 + y_2^2)K_{1,n}K_{1,n-1} + jy_2(1 - 2K_{1,n}K_{2,n-1}) - K_{2,n}K_{2,n-1}}{y_1^2K_{1,n}^2 - (K_{2,n} + jy_2K_{1,n})^2}. \tag{844}$$

For a purely real y the resistance (843) becomes

$$\frac{\mathcal{R}}{R} = \frac{y}{y^2K_{1,n}^2 - K_{2,n}^2} \tag{843a}$$

Specifically, when the right-hand network in Fig. 146 is a properly terminated mid-series section of a constant-k filter, then

$$\frac{R}{y} = R\sqrt{1 - x_k^2},$$

so that for the region $-1 \leq x_k \leq 1$ we have by (843a)

$$\frac{\mathcal{R}}{R} = \frac{\sqrt{1 - x_k^2}}{K_{1,n}^2 - K_{2,n}^2(1 - x_k^2)}. \tag{843b}$$

The denominator of this expression is evidently a polynomial in x_k^2. Since n is even, $K_{1,n}$ contains the term unity as well as products of all even numbers of elements up to n, while $K_{2,n}$ contains terms involving all odd numbers of elements from one to $n - 1$. Hence the polynomial in x_k^2 contains all powers from zero to n, so that we have

$$\frac{\mathcal{R}}{R} = \frac{\sqrt{1 - x_k^2}}{1 + A_1x_k^2 + A_2x_k^4 + \cdots + A_nx_k^{2n}}. \tag{845}$$

The structure of Fig. 146, therefore, is capable of correcting the resistance component of the mid-series constant-k characteristic impedance to the extent that the expression (845) is capable of approximating unity over the range $-1 \leq x_k \leq 1$. The ability of the function (845) to do this increases with the number of arbitrary coefficients $A_1 \ldots A_n$. By a proper choice of these coefficients, this function may be made to behave much like the characteristic impedances of the higher m-derived types.

The problem of determining the proper values for the polynomial coefficients so that the function will maintain a uniform maximum tolerance over the transmission range is purely an algebraic one. It is possible here to plot curves of tolerance and coverage versus the various coefficients as we did for Zobel's m and m'.

Bode suggests two methods. One[1] is to expand the radical in a power series, breaking off with the $2n$th degree in x_k, and then equate coefficients of like powers. The resulting functions vary monotonically, i.e., show no maxima or minima. This scheme does not utilize the ability of the function to the fullest extent. The second method[2] proceeds by expanding the radical in terms of Legendrian harmonics (also called spherical polynomials) and after substituting these polynomials

[1] *Loc. cit.*, p. 21.

[2] *Loc. cit.*, p. 23.

and sorting out the various powers of x_k the coefficients are again equated to the corresponding A's. This determination results in a function (845) which oscillates about the value unity and gives rise to the best "least squares" approximation.

The expansion of the radical reads[1]

$$\sqrt{1-x^2} = \frac{\pi}{2}\left\{\frac{1}{2}P_0(x) - 5\left(\frac{1}{4}\right)\left(\frac{1}{2}\right)^2 P_2(x) - 9\left(\frac{3}{6}\right)\left(\frac{1}{2\cdot 4}\right)^2 P_4(x) - 13\left(\frac{5}{8}\right)\left(\frac{1\cdot 3}{2\cdot 4\cdot 6}\right)^2 P_6(x) - \cdots\right\},$$

and the spherical polynomials are given by[2]

$$\begin{aligned} P_0(x) &= 1 \\ P_2(x) &= \tfrac{1}{2}(3x^2 - 1) \\ P_4(x) &= \tfrac{1}{8}(35x^4 - 30x^2 + 3) \\ P_6(x) &= \tfrac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5). \end{aligned}$$

For various values of n, Bode thus finds the following A's and corresponding a's:

n	A_1	A_2	A_3	a_1	a_2	a_3
1	−0 7142	0	0	0.5546	0	0
2	−0 3236	−0 4885	0	0 8986	1 593	0
3	−0 0461	+0 4958	−0 7162	0 9597	1 924	1 565

The resulting reactance, which may be determined from (844), must then be annulled by means of a series reactance network. This is designed quite readily by a cut-and-try process utilizing Foster's theorem as already discussed.

Fundamentally the method of impedance correction described here could be used to make a given impedance look more or less like any other impedance. It, therefore, is of importance wherever it is desired to make a given impedance simulate some other desired function. The method is, of course, most readily carried out in such cases where the forms of y_1 and y_2 in (843) and (844) are such as to reduce these to quotients of simple polynomials.

13. Other conventional filter types. The stipulation that the component reactances z_1 and z_2 of a uniform ladder structure shall be inverse with respect to a constant (which is basic for the constant-k type) limits the variety of possible results obtainable from such a structure. By dropping this requirement, a number of additional selective character-

[1] Byerly, "Fourier Series and Spherical Harmonics," Ginn & Co., p. 184.

[2] Byerly, p. 151.

istics are made available in the band-pass, band-elimination, and higher filter classes. Since the process of m-derivation is generally applicable to networks of the uniform ladder type, such additional conventional filter prototypes may be used in the formation of composite structures by means of the same methods as are applied to the constant-k.

The design process for these prototypes is based upon the principles discussed in section 1 which hold for all symmetrical T- or Π-networks. For example, if, in the constant-k band-pass network, we omit one element in either the series or shunt reactance, the resulting structure is still a band-pass filter although its attenuation function is somewhat altered. Suppose we have

$$z_1 = \frac{1 - L_1C_1\omega^2}{jC_1\omega}; \; z_2 = jL_2\omega,$$

i.e., let the shunt arm be given by a coil alone instead of by an anti-resonant component, then

$$-\frac{z_1}{4z_2} = x^2 = \frac{1 - L_1C_1\omega^2}{4L_2C_1\omega^2},$$

and

$$x = \pm\sqrt{\frac{1 - L_1C_1\omega^2}{4L_2C_1\omega^2}} = \xi + j\eta.$$

According to (679), therefore, we have for the lower attenuation range

$$\left.\begin{array}{l} 0 < \omega < \dfrac{1}{\sqrt{(L_1 + 4L_2)C_1}} \\ -\infty < \xi < -1; \; \eta = 0 \\ \infty > \gamma_1 > 0; \; \gamma_2 = -\pi \end{array}\right\},$$

for the transmission range

$$\left.\begin{array}{l} \dfrac{1}{\sqrt{(L_1 + 4L_2)C_1}} < \omega < \dfrac{1}{\sqrt{L_1C_1}} \\ -1 < \xi < 0; \; \eta = 0 \\ \gamma_1 = 0; \; -\pi < \gamma_2 < 0 \end{array}\right\},$$

and for the upper attenuation range

$$\left.\begin{array}{l} \dfrac{1}{\sqrt{L_1C_1}} < \omega < \infty \\ \xi = 0; \; 0 < \eta < \sqrt{\dfrac{L_1}{4L_2}} \\ \gamma_2 = 0; \; 0 < \gamma_1 < 2\sinh^{-1}\sqrt{\dfrac{L_1}{4L_2}} \end{array}\right\}.$$

We note here a dissymmetry as compared with the variation of these quantities in the constant-k filter. Instead of passing through the values -1 to $+1$, x varies through -1 to 0 and then becomes imaginary. The cut-off frequencies correspond to $x = -1$ and $x = 0$, respectively. The attenuation characteristic is dissymmetrical. It is infinite at zero frequency (like that for the constant-k) but approaches a finite value at infinity. When the width of the pass band is small compared with the mean frequency, this dissymmetry is much less evident. The reader should recognize that this type of band-pass filter is equivalent to a pair of identical tuned mutually coupled circuits as commonly used in radio work.

The characteristic impedance of this prototype has the same functional form in the pass band as that for the mid-series constant-k section as may be seen by carrying out the proper substitutions. When the band width is narrow compared with its distance from the origin, the analysis may be simplified by assuming for the reactances the approximate expressions

$$z_1 = 2jL_1(\omega - \omega_2);\ z_2 = jL_2\omega_m,$$

where

$$\omega_2 = \frac{1}{\sqrt{L_1C_1}};\ \omega_1 = \left(\frac{L_1 - L_2}{L_1 + L_2}\right)\omega_2$$

are the cut-off frequencies, and ω_m their arithmetic mean. z_2 is thus assumed constant and equal to its value at the mid-band frequency. For the vicinity of the pass band this is a legitimate procedure. The band width is then given by

$$w = \omega_2 - \omega_1 = \frac{2L_2}{L_1 + L_2} \cdot \frac{1}{\sqrt{L_1C_1}},$$

and the characteristic impedance becomes

$$Z_T = L_1\sqrt{(\omega_2 - \omega)(\omega - \omega_1)},$$

which, at the mean frequency ω_m, has the maximum value

$$Z_{T\max} = \frac{L_1w}{2} = \frac{L_1L_2}{L_1 + L_2} \cdot \frac{1}{\sqrt{L_1C_1}}.$$

For the determination of the propagation function we have

$$\frac{z_1}{4z_2} = -x^2 = \frac{\omega - \omega_2}{w},$$

$$x = \sqrt{\frac{\omega_2 - \omega}{w}}.$$

For the lower attenuation range this is real and (679) gives

$$\cosh^2 \frac{\gamma_1}{2} = \frac{\omega_2 - \omega}{w},$$

while for the upper attenuation range x is imaginary so that

$$\sinh^2 \frac{\gamma_1}{2} = \frac{\omega - \omega_2}{w}.$$

Applying formulae for hyperbolic functions of whole in terms of half angles, these are respectively equivalent to

$$\left.\begin{aligned} \cosh\gamma_1 &= \frac{2(\omega_m - \omega)}{w} \\ \cosh\gamma_1 &= \frac{2(\omega - \omega_m)}{w} \end{aligned}\right\},$$

which show that the approximations made here lead to a symmetrical characteristic. Making the same approximations for the constant-k band-pass filter, the relations (686) and (829a) (pages 304 and 363) show that we get

$$\left.\begin{aligned} \cosh\frac{\gamma_{1k}}{2} &= \frac{2(\omega_m - \omega)}{w} \\ \cosh\frac{\gamma_{1k}}{2} &= \frac{2(\omega - \omega_m)}{w} \end{aligned}\right\}$$

respectively. This shows that the prototype considered here has a considerably lower attenuation characteristic. At a frequency $(\omega - \omega_m)$ the constant-k has *twice* the attenuation that the new prototype has at this same frequency.

Note also that if the filter is terminated in a resistance equal to the maximum of Z_T in the pass band, i.e., if we put

$$R = Z_{T\ \max},$$

then from above we have

$$L_1 = \frac{2R}{w};\ C_1 = \frac{w}{2R\omega_2^2};\ L_2 = \frac{R}{\omega_m}.$$

The first two of these are practically the same as for the constant-k filter, as given by (735), page 319.[1]

[1] Note that for the constant-k filter $L_2/L_1 = (w/2\omega_o)^2$ whereas for the present prototype $L_2/L_1 = w/2\omega_o$. In designing a filter for radio frequencies, for example, the ratio $w/2\omega_o$ is very small (about 1/200). If we attempt a constant-k design, the inductance L_2 becomes so small compared with L_1 that the structure is impractical. For the prototype discussed here, this situation is less unfavorable. This accounts for the fact that tuned coupled circuits (or their equivalents) are used for band selection at high frequencies in spite of the fact that they afford relatively poor selective characteristics.

This simple example should suffice to illustrate the general characteristics obtainable with prototypes of this sort. The reader may investigate other element combinations as practice problems.

PROBLEMS TO CHAPTER IX

9–1. Construct a chart of the type shown in Fig. 109 with sufficient attenuation and phase loci so that it may be used for ordinary graphical work. The attenuation range of the chart should extend to 50 db.

9–2. A constant-k low-pass filter is designed for $R = 1000$ and $\omega_c = 2\pi \cdot 2000$. The R/L ratio of the coils is assumed constant and equal to 100. Determine and plot the attenuation and phase functions (γ_1 and γ_2) for a single section over the range 0–4000 cycles per second, and compare with the corresponding non-dissipative values. Assume negligible loss in the condenser.

9–3. A constant-k filter is to have an external transmission range $0 < \omega < \omega_{c1}$ and also an internal transmission range $\omega_{c2} < \omega < \omega_{c3}$. Sketch the structures for z_{1k} and z_{2k} which are potentially able to meet these specifications, and sketch a resulting T-section. Write the analytic expressions for these component reactances and the equations from which the parameter values may be uniquely determined in terms of the nominal resistance R. Sketch the functions z_{1k} and $4z_{2k}$ vs. ω and show that the general form of these curves satisfies eqs. (682) and (690).

9–4. Make plots corresponding to those shown in Figs. 110 and 111 over a range $-6 < x_k < 6$ with sufficient care so that they may be used with a fair degree of precision for subsequent graphical work.

9–5. For the filter of Problem 9–3 make sketches vs. frequency of the functions γ_1, γ_2, W_{1k}, and W_{2k} over a range which extends sufficiently beyond the last boundary to illustrate the general character of these functions.

9–6. Repeat the sketches called for in Problems 9–3 and 9–5 for the constant-k **L.P.**, **H.P.**, **B.P.**, and **B.E.** classes.

9–7. Consider the constant-k **B.P.** structure modified by (*a*) omitting L_1; (*b*) omitting C_1; (*c*) omitting L_2; (*d*) omitting C_2. In each case sketch the reactance functions z_1 and $4z_2$, and show that the filter is still of the **B.P.** class although no longer of the constant-k type. Sketch the attenuation, phase, and characteristic impedance functions for this modified filter with sufficient care to show the essential differences between these functions and those for the corresponding constant-k type.

9–8. Design a constant-k **B.P.** filter and one of the modified type discussed in Problem 9–7 and in section 13 for cut-offs at 1000 and 2000 cycles per second and a nominal characteristic impedance of 1000 ohms. Plot the attenuation loss for each design over the range 0–4000 cycles per second, and compare these for the two filter types. Repeat for the cut-off frequencies 175,000 and 185,000 cycles per second, plotting in this case over the range 140,000–220,000 cycles per second. For this second design show that the approximations discussed in the latter part of section 13 are justified in the modified type of filter and that for the constant-k type the geometric symmetry is very nearly identical with an arithmetic symmetry about the mid-band frequency.

9–9. A low-pass filter is to work out of 500 and into 1000 ohms and have a cut-off at 3000 cycles per second. At 3500 cycles per second the attenuation is to be at least 40 db. Carry out a design on the basis of a cascade of half-sections according to Fig. 114. Neglect reflection effects in determining the loss in the structure.

9–10. Determine the reflection and interaction losses and thus the insertion loss for the design of Problem 9–9, and plot over the range 3000–6000 cycles per second

9–11. By incorporating the equivalent transformer circuits of Figs. 42 and 43 in the design of a half-section of a constant-k band-pass filter and utilizing the methods discussed in section 4, it is possible to design a filter employing mutual inductive coupling and realizing appreciable ratios in impedance levels between the input and output ends. On the basis of these ideas carry through a design for a band-pass filter which will work out of 500 ohms and into 5000 ohms for a band width of 10,000 cycles per second and a mid-band frequency of 10^6 cycles per second utilizing a single half-section. Calculate and plot the attenuation function over a range extending from 20,000 cycles below the lower cut-off to 20,000 cycles above the upper cut-off, and compare the results with those obtained for a pair of tuned coupled circuits giving the same range of frequency selection designed according to the method outlined in Problems 8–4 and 8–5. The latter is, of course, a symmetrical network, but it is interesting to compare the resulting losses only. The design proposed in this problem offers a variety of possibilities according to the manner in which the equivalent transformer is oriented with regard to the filter half-section. Choose the arrangement which leads to the most favorable design, and discuss this from the standpoint of practicability.

9–12. A low-pass filter is to be designed for the following specifications: The theoretical cut-off should lie at 3000 cycles per second; at 3500 cycles per second the attenuation shall be at least 40 db, and it shall not drop below this value at higher frequencies; the characteristic impedance shall have a nominal value of 800 ohms and may be of the form obtainable from a constant-k design, i.e., this form is sufficiently good. The design may, therefore, be met by cascading a sufficient number of constant-k sections, although a composite type which also utilizes the sharper cut-off properties of the derived types may lead to better overall economy. Investigate all possible means for meeting the stated requirements, and discuss from the standpoint of optimum economy. Neglect reflection and interaction losses.

9–13. For the low-pass filter of Problem 9–12 let the specification for the characteristic impedance be that it shall not deviate from a mean value of 800 ohms by more than 3 per cent over 86 per cent of the theoretical transmission range. Reconsider the design under these conditions.

9–14. For the optimum design in Problem 9–12 determine the reflection and interaction losses in the attenuation range, and see how their effect influences the adequacy of the resulting structure.

9–15. Repeat Problem 9–14 for the design specified in Problem 9–13.

9–16. Reconsider the design in Problem 9–12 with the altered attenuation requirement that the loss shall be at least 50 db for the range 3500–4500 cycles per second but need not exceed 30 db for frequencies higher than 5000 cycles per second.

9–17. Calculate and plot the time delay for the filter design in Problem 9–12 consisting of constant-k sections alone. Repeat for a composite design meeting the same specifications.

9–18. A composite filter consisting of a constant-k section with symmetrically placed half-sections of the M-type is designed for the internal transmission range 3000–6000 cycles per second and the external range 10,000–∞. Sketch a structure capable of meeting these requirements. If the M-type is designed for $m = 0.6$, sketch the overall attenuation and phase functions ($\gamma_{1k} + \gamma_{1km}$ and $\gamma_{2k} + \gamma_{2km}$) and the mid-series impedance W_{1km} over the range 0–15,000 cycles per second. What are the minimum values of attenuation? Set down the equations from which you

may determine the frequencies at which the minimum and infinite values of attenuation occur.

9–19. A composite band-elimination filter consisting of a constant-k type T-section with series M-type half-sections is designed for $\omega_{c1} = 10{,}000$; $\omega_{c2} = 20{,}000$; $m = 0.6$. Determine:

(a) The frequencies at which infinite attenuation occurs.
(b) " " " " minimum " "
(c) " " " " W_{2km} is zero or infinite.
(d) The minimum attenuation values.

9–20. For the filter of Problem 9–19 determine and plot the insertion loss over the attenuation band and compare with the attenuation loss, assuming the filter to be terminated in its nominal resistance value of R ohms at either end. Show that the insertion loss is the same if the constant-k section is mid-shunt and the M-type half-sections are shunt derived.

9–21. A composite band-pass filter consisting of a constant-k section with M and MM'-type half-sections symmetrically placed should have a characteristic impedance which does not deviate from a mean value of 600 ohms by more than 1.5 per cent over 96 per cent of the theoretical band. The theoretical cut-off frequencies should lie at 10,000 and 13,000 cycles per second. Determine the structure and its parameter values. Plot the attenuation and phase functions over the range 0–20,000 cycles per second. Determine the minimum values of attenuation and the frequencies at which they occur. Without carrying through an exact calculation, indicate approximately the quantitative effect which the reflection and interaction losses have, and sketch the resulting insertion loss.

9–22. Design constant-k low-pass and high-pass filters for a common cut-off of 3500 cycles per second, and join them in parallel on their input sides by means of fractional series terminations. Supply the output ends individually with mid-series derived half-sections, and assume the whole to work out of and into the nominal resistance R. Calculate and plot the resulting attenuation and phase functions.

9–23. A pair of band-pass filters have their theoretical bands located in the ranges 10,000–13,000 and 13,000–16,000 cycles per second, respectively, and are to be connected in parallel on their input ends by supplying each with a fractional series termination with $a = 0.6$ and annulling the resulting susceptance over their combined transmission range. Design a susceptance annulling network consisting of two resonant components in parallel, and plot the residual susceptance.

9–24. Repeat the design and plots of Problem 9–21, using Bode's method of impedance correction in place of Zobel's method of repeated m-derivations for meeting the required terminal conditions.

CHAPTER X

FILTER DESIGN METHODS BASED UPON THE LATTICE STRUCTURE

1. Introductory remarks. The chief disadvantage of the methods of filter design discussed in the previous chapter lies in their lack of flexibility in meeting arbitrary requirements regarding the behavior of the propagation and characteristic impedance functions. Each additional step in the development of the composite structure carries with it a partial improvement in both functions; it never affords an independent adjustment in either one of these. By some happy chance an improvement in the propagation function carries with it a potential improvement in the final characteristic impedance, while an improvement in the latter definitely effects an improvement in the former. Each successive derivation introduces an additional design parameter such as m, m', m'', etc., but the choice of its numerical value is always based upon a compromise. This property is inherent in the symmetrical T- and Π-structures upon which the previous design methods are based.

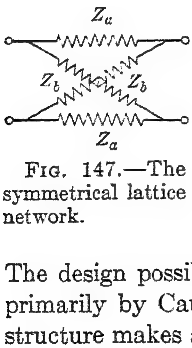


FIG. 147.—The symmetrical lattice network.

It is possible to avoid such restrictions by utilizing the symmetrical lattice or bridge as a basic structure. The design possibilities afforded by the latter have been investigated primarily by Cauer[1] and Bode,[2] whose studies show that this form of structure makes available a much more flexible and general method for the design of selective networks. In addition to this, the type of reasoning called forth by this more general method aids in clarifying and correlating the steps involved in the procedure already discussed.

The reason for these advantages of the lattice structure may be appreciated in a general way by referring to the corresponding expressions for the propagation and characteristic impedance functions. Fig. 147

[1] W. Cauer, "Die Siebschaltungen der Fernmeldtechnik," Zeitsch. f. angew. Math. u. Mech., vol. 10, pp 425–433, 1930; "Siebschaltungen," published by V.D.I. Verlag G.m.b.H., Berlin, 1931; "New Theory and Design of Wave Filters," Physics, Vol. 2, No. 4, p 242, April, 1932.

[2] H. W. Bode, U. S. Patent No. 1,828,454, Oct. 20, 1931; "A General Theory of Electric Wave Filters," Jour. of Math. and Phys., Vol. 13, No. 3, pp. 275–362, November, 1934; H. W. Bode and R. L. Dietzold, "Ideal Wave Filters," B.S.T.J. Vol. XIV, pp. 215–252, April, 1935.

illustrates the structure. According to equations (430) and (431), page 179, we have for these functions respectively

$$\left.\begin{aligned} \gamma &= \ln\left(\frac{\sqrt{\dfrac{Z_b}{Z_a}}+1}{\sqrt{\dfrac{Z_b}{Z_a}}-1}\right) = 2\tanh^{-1}\sqrt{\frac{Z_a}{Z_b}} \\ &\text{and} \\ Z_0 &= \sqrt{Z_bZ_a} \end{aligned}\right\}. \tag{846}$$

The choice of functions for γ and Z_0 thus fixes only the quotient and product $(Z_b/Z_a)^{\frac{1}{2}}$ and $(Z_bZ_a)^{\frac{1}{2}}$, from which we get by multiplication and division

$$\left.\begin{aligned} Z_b &= (Z_bZ_a)^{\frac{1}{2}} \cdot (Z_b/Z_a)^{\frac{1}{2}} \\ Z_a &= (Z_bZ_a)^{\frac{1}{2}} \,/\, (Z_b/Z_a)^{\frac{1}{2}} \end{aligned}\right\}. \tag{846a}$$

Only such functions may, of course, be chosen as will lead to physically realizable impedances according to (846a), but it is evident that the functions for γ and Z_0 may here be chosen *independently.* This fact provides more freedom in meeting the individual requirements called for by a specific design problem. Thus, for example, we may wish to have a high degree of matching at the terminals of the filter and not require an abrupt cut-off or a high attenuation beyond this boundary. With the previous design method we were more or less obliged to accept one of these characteristics with the other whether this was specifically required or not.

In thus making the lattice structure basic for the design procedure it does not necessarily follow that this form must be used in the practical realization of the filter. There are several reasons why the lattice is less desirable than the ladder form in carrying out the construction of a filter. A fairly obvious reason is the fact that in the lattice each component reactance enters twice, whereas in the T- or Π-structures (which form the basis of previous design procedures) one of the reactances enters only once. This would seem to indicate a loss in economy in using the lattice. Where the practical difficulties presented by the introduction of ideal transformers may be overcome, this objection is obviated by the use of one of the equivalent structures discussed in section 7 of Chapter IV.

It is important to note, however, that economy in the practical construction of a filter is not measured merely in terms of the total number of coils and condensers embodied in a given physical form. A frequently more important point is the tolerance to which the inductance

and capacitance values must be held in order to realize the theoretically predicted behavior with a sufficient degree of approximation. In this respect the lattice structure (being a Wheatstone bridge) finds itself at a marked disadvantage as compared with the ladder form. It is for this reason that a lattice design is converted into an equivalent ladder development whenever possible, even though the latter may involve a larger total of coils and condensers. In fact, the allowable tolerances in the various inductances and capacitances imposed by a lattice frequently make its physical realization impossible.

We shall see in a later section, where these matters will be discussed in more detail, that a complete ladder development of the lattice is not always possible. In such a case the undeveloped portion which is still in the lattice form may sometimes be split into a cascade of simpler lattices for which the tolerances on the elements are not too severe to prohibit their realization. The ladder developments arrived at in this process are not uniform ladder structures, of course; otherwise the design process to be taken up here should have been given by the methods discussed earlier. On the other hand it is true, as Bode points out,[1] that the end-results obtainable can be had by other lines of approach which do not bring the lattice per se into the design picture. In a large measure the introduction of the lattice is merely a design artifice which, at the present state of the art, is an effective means toward attaining the desired end.

A further reason for choosing the lattice as a basic structure is brought out by the following reasoning. In section 10 of Chapter V we discussed the general conditions to be fulfilled by a set of three reactances in order that they may represent the z-system of a physical network. These conditions are expressed by equation (519) which, for the symmetrical case, becomes

$$(k_{11}^2 - k_{12}^2) = (k_{11} + k_{12})(k_{11} - k_{12}) \geq 0, \tag{847}$$

where the superscripts are dropped because they are of no consequence in our present argument. Since k_{11} (the residue of z_{11}) must be positive, this condition splits into the following two

$$\left.\begin{aligned} (k_{11} + k_{12}) \geq 0 \\ (k_{11} - k_{12}) \geq 0 \end{aligned}\right\}. \tag{847a}$$

If we now recall that for the symmetrical lattice

$$\left.\begin{aligned} z_{11} = \tfrac{1}{2}(Z_b + Z_a) \\ z_{12} = \tfrac{1}{2}(Z_b - Z_a) \end{aligned}\right\},$$

[1] See his *Substitution method of introducing impedance controlling factors* and his discussion of *H-derived networks*, "A General Theory . . . " loc. cit., pp. 348–362.

so that

$$\left.\begin{aligned} Z_b &= z_{11} + z_{12} \\ Z_a &= z_{11} - z_{12} \end{aligned}\right\},$$

we recognize that $(k_{11} + k_{12})$ is the residue of Z_b and $(k_{11} - k_{12})$ that of Z_a. *The necessary and sufficient conditions for the physical realization of an arbitrary symmetrical reactive four-terminal network, therefore, become identical to those for the physical realization of the component reactances of a symmetrical lattice.* This means that *if a symmetrical reactive four-terminal network is realizable at all, it is realizable in the lattice form.* In studying the lattice we are, therefore, studying the most general form of symmetrical network.[1] This is an important advantage over the T- or Π-structures previously used.

2. Conditions for transmission and attenuation. In the following treatment we shall again neglect dissipation and consider the components Z_a and Z_b as pure reactances.

The discussion of design relationships becomes somewhat simpler if instead of Z_a and Z_b we use the so-called *normalized reactances*

$$\left.\begin{aligned} z_a &= \frac{Z_a}{R} \\ z_b &= \frac{Z_b}{R} \end{aligned}\right\}, \tag{848}$$

where R is again the nominal value[2] of the characteristic impedance and hence a positive real constant. Thus

$$\frac{Z_0}{R} = z_0 = \sqrt{z_b z_a} \tag{849}$$

becomes the *normalized characteristic impedance*, the nominal value of which is unity. The propagation function, according to (846), retains the same form whether written in terms of Z_a and Z_b or z_a and z_b.

[1] This statement holds true for the dissipative case also. There the condition replacing (847), see footnote 2 on p. 217, is $(r_{11}^2 - r_{12}^2) = (r_{11} + r_{12})\ (r_{11} - r_{12}) \geq 0$ for $\gamma \geq 0$, $(\lambda = \gamma + j\omega)$, r_{11} and r_{12} being the real parts of $z_{11}(\lambda)$ and $z_{12}(\lambda)$. Since $r_{11} \geq 0$ for $\gamma \geq 0$ is necessary for the realizability of z_{11}, we have $(r_{11} + r_{12}) \geq 0$ and $(r_{11} - r_{12}) \geq 0$ for $\gamma \geq 0$. These, however, are again the conditions for the realizability of Z_b and Z_a. An alternative proof, restricted to networks which may be bisected so as to form symmetrical halves, is given by A. C. Bartlett in his paper "An Extension of a Property of Artificial Lines," Phil. Mag., Nov. 1927, pp. 902–907 (see particularly his corollary 1, p. 906). A general proof was first given by W. Cauer in his paper "Ueber die Variabeln eines passiven Vierpols," Sitzungsberichte d. Preuss. Akad. d. Wiss., Dec. 1927; and an alternative simpler proof in his paper "Vierpole" (E.N.T. 6, No. 7, 1929) on page 279.

[2] Here the nominal value is understood to be that resistance which the filter is to work into and out of.

From the first relation (846) we may determine the conditions for the realization of transmission and attenuation bands by inspection. Thus if (z_b/z_a) is negative, $(z_b/z_a)^{\frac{1}{2}}$ is a pure imaginary. The argument of the logarithm is then in the form of the quotient of a complex number and its negative conjugate. This quotient has the magnitude unity, and hence γ is a pure imaginary. This, of course, is also evident from the inverse hyperbolic tangent. The condition for the realization of a transmission range is, therefore, that

$$(z_b/z_a) < 0, \tag{850}$$

which is the same as to say that z_a and z_b shall have opposite signs. This condition makes the product $(z_b z_a)$ positive (since z_b and z_a are pure imaginaries), and hence the characteristic impedance becomes a real quantity, but not necessarily a real constant. It will be a real constant if z_b and z_a are not only of opposite sign but inverse functions with respect to a constant. In order for the characteristic impedance to be constant and equal to its nominal value throughout a transmission range, the normalized component reactances must satisfy the relation

$$z_a z_b = 1. \tag{851}$$

Since this condition insures (850), it becomes the definition for the *ideal* transmission range.

An attenuation range, on the other hand, occurs wherever z_b and z_a have the same sign, as may also be seen from (846). Furthermore, it is evident that the attenuation will be infinite when $z_b = z_a$, or if

$$\frac{z_b}{z_a} = 1. \tag{852}$$

This, therefore, is the definition for the *ideal* attenuation range. In such a range Z_0, of course, is imaginary and behaves like a physical reactance.

In the design of a filter of this type, the object is to realize, as nearly as possible or as nearly as is required, the ideal conditions (851) and (852) in the prescribed transmission and attenuation ranges. The exact realization of these conditions, of course, is impossible since this would mean that the component reactances would have to change suddenly and completely from the behavior (851) to that defined by (852), or vice versa, at each boundary. It is not hard to appreciate that this is too much to expect from ordinary physical reactances. We may expect, however, that the transition from the one behavior to the other can become reasonably abrupt and complete by making the reactive networks for z_b and z_a sufficiently extensive and by properly

choosing their resonant and anti-resonant frequencies. How this may be accomplished in order to meet prescribed tolerances and limits for both the propagation and characteristic impedance functions is the subject of this method of filter design.

3. The approximation problem. For the following discussion it is convenient to introduce the function

$$y_0 = \sqrt{\frac{Z_b}{Z_a}} = \sqrt{\frac{z_b}{z_a}}, \tag{853}$$

which plays the same part for the lattice filter that y_k, given by equation (749), page 325, does for the constant-k type. In the transmission range y_0 is a pure imaginary quantity; in the attenuation range it is real. There it is characteristically indicative of the attenuation, the latter becoming infinite for $y_0 = 1$. The function y_0, therefore, will be referred to as the **index function.**

Our present problem is one of determining reactances z_b and z_a such that the functions $y_0 = (z_b/z_a)^{\frac{1}{2}}$ and $z_0 = (z_b z_a)^{\frac{1}{2}}$ approximate unity in the attenuation and transmission ranges, respectively, in a prescribed manner. In order to illustrate in some detail the general form which this problem takes, let us consider specifically the case of a low-pass filter. Fig. 148 illustrates a possible distribution of zeros and poles for a pair of reactances z_b and z_a. For the range $0 < \omega < \omega_1$ these have opposite signs; for $\omega_1 < \omega < \infty$ they have the same sign. Thus these ranges become the transmission and attenuation ranges, respectively. The frequency ω_1 marks the boundary, and hence is the cut-off frequency. The total number of zeros and poles is arbitrarily chosen, and infinity is indicated as a finite point for convenience.

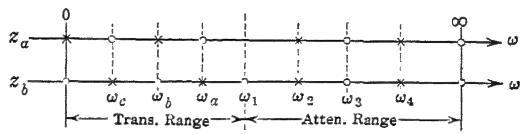


FIG. 148.—Possible distribution of critical frequencies for the component reactances of a low-pass lattice-type filter.

In factored form the reactances are given by

$$\left.\begin{aligned} z_a &= \frac{\mu_a(\omega_c^2 - \omega^2)(\omega_a^2 - \omega^2)(\omega_3^2 - \omega^2)}{j\omega(\omega_b^2 - \omega^2)(\omega_2^2 - \omega^2)(\omega_4^2 - \omega^2)} \\ z_b &= \frac{\mu_b j\omega(\omega_b^2 - \omega^2)(\omega_1^2 - \omega^2)(\omega_3^2 - \omega^2)}{(\omega_c^2 - \omega^2)(\omega_a^2 - \omega^2)(\omega_2^2 - \omega^2)(\omega_4^2 - \omega^2)} \end{aligned}\right\}, \tag{854}$$

where the notation for the critical frequencies is as indicated in the figure, and μ_a and μ_b are positive real factors. In the transmission range

the zeros and poles of the two functions replace each other; in the attenuation range they are coincident. The boundary frequency ω_1 is a zero of z_b which is not contained in z_a. We could just as well have allotted this critical frequency to z_a, or chosen a pole instead of a zero at the boundary, the significant point being that this critical frequency is contained in *only one* of the component reactances.

From (854) we then find

$$y_0 = \sqrt{\frac{z_b}{z_a}} = \sqrt{\frac{\mu_b}{\mu_a}} \cdot \frac{j\omega(\omega_b^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{(\omega_c^2 - \omega^2)(\omega_a^2 - \omega^2)}, \tag{855}$$

and

$$z_0 = \sqrt{z_b z_a} = \sqrt{\mu_b \mu_a} \cdot \frac{\sqrt{\omega_1^2 - \omega^2}(\omega_3^2 - \omega^2)}{(\omega_2^2 - \omega^2)(\omega_4^2 - \omega^2)}. \tag{856}$$

The functions y_0 and z_0 are obviously real in the attenuation and transmission ranges, respectively. In the transmission range y_0 is imaginary; in the attenuation range z_0 is imaginary. There these functions have simple zeros and poles separating each other and positive slope; i.e., they behave like physical reactances. At the cut-off frequency ω_1, their behavior is governed by the irrational factor $\sqrt{\omega_1^2 - \omega^2}$ which is contained in both and is called the **cut-off factor.**

The significant point is that y_0 and z_0 do not contain the same set of design parameters. Thus the behavior of y_0 is controlled by the parameters

$$(\mu_b/\mu_a)^{\frac{1}{2}};\ \omega_c;\ \omega_b;\ \omega_a,$$

whereas z_0 is controlled by

$$(\mu_b\mu_a)^{\frac{1}{2}};\ \omega_2;\ \omega_3;\ \omega_4.$$

The number of design parameters in y_0 and z_0 is independently determined by the number of zeros and poles chosen within and without the transmission range, respectively. Hence either function may be given any degree of complexity independently of the other. Their general forms, of course, are limited to those given by (855) and (856). The approximation problem usually consists in determining the parameters in these functions so that they become as nearly equal to unity in their respective real ranges as their degree of complexity will allow. Each additional factor makes possible a closer approximation, so that the design of the filter may be made as nearly ideal as desired by adding a sufficient number of zeros and poles to z_b and z_a.

The manner in which the solution to the approximation problem is carried out depends to a large extent upon the requirements of the

design. Thus a specific design may call for a particular manner in which the functions y_0 and z_0 are to approximate unity within their respective real ranges. In this connection it must be remembered also that the behavior of y_0 not only influences the attenuation function in the attenuation range *but also the phase function in the transmission range.* Fixing the manner in which y_0 approximates unity in the attenuation range determines not only the attenuation function in that range but also the phase function in the pass band. If it is essential to have the phase function in the pass band as linear as possible, the behavior of y_0 in the attenuation range must be determined with this requirement as the governing factor.

Whatever the situation may be in a specific case, the parameter determination, of course, may always be carried out by means of a cut-and-try process. Since this usually consumes a large amount of time, it is well to have available some "ready-made" parameter determinations which may be applied directly to certain average cases or used as first approximations in special ones where a more extensive cut-and-try method is necessary. The general character of the functions in two such determinations is illustrated by curves A and B of Fig. 149. In the low-pass case z_0 is considered as a function of ω/ω_1 and y_0 of ω_1/ω. For the high-pass these functions are interchanged. The curves correspond to the functions (855) and (856) in which the cut-off factor appears in the numerator of both. If this factor occurs in the denominator, the curve goes to infinity at $\omega = \omega_1$ instead of becoming zero there, but the general character is otherwise unchanged.

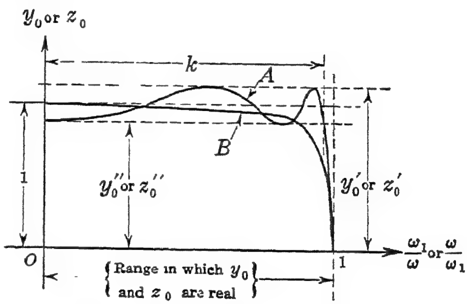


FIG. 149.—Oscillatory (Cauer) and monotonic (Bode) forms of behavior for the functions y_0 and z_0 in their approximation ranges.

Curve A illustrates an *oscillatory* form of approximation. In the case of y_0, the points at which the curve crosses unity correspond to infinite maxima for the attenuation function, while the points of maximum deviation from unity correspond to minima in the attenuation function. If the maximum *positive* deviations are all equal and given by $y_0 = y_0'$ while all the maximum *negative* deviations are also equal and given by $y_0 = y_0''$ such that

$$y_0'y_0'' = 1, \tag{857}$$

then all the minimum values of the attenuation function become equal. This is easily seen from (846) and (853) which give

$$\gamma_1 = \ln\left|\frac{y_0 + 1}{y_0 - 1}\right|. \tag{858}$$

Reciprocal values of y_0 evidently give rise to the same attenuation.

Since the reflection coefficient is given by

$$r = \frac{1 - z_0}{1 + z_0}, \tag{859}$$

it follows that the type of approximation shown by curve A when utilized for z_0 with maxima equal to z_0' and minima equal to z_0'' such that

$$z_0' z_0'' = 1, \tag{860}$$

leads to *equal* maximum deviations from zero on the part of the magnitude of the reflection coefficient. Functions y_0 and z_0 which behave in this way are said to approximate unity in the *Tschebyscheff manner.* This is the type of behavior which Cauer chooses in his method of filter design and to which his method of parameter determination applies. The details of the latter will be given in a later section.

Although this oscillatory form of approximation seems to meet the average design requirements most directly, it possesses one major defect, namely, that it leads in the case of y_0 to an undesirable phase characteristic in the transmission range. In this respect the present situation is very similar to that encountered in the Zobel type of composite filter. Here it will be recalled that the phase characteristic of the M-type is more concave than that of the constant-k, as may be seen by comparing Figs. 120 and 110 (pages 329 and 304), respectively. For the higher order derived types, the phase characteristic becomes worse instead of better. Where a linear phase characteristic in the transmission range is essential, the Cauer method of parameter determination is not desirable, therefore, unless the phase is subsequently corrected by an additional network.

Bode has shown[1] that a method of parameter determination for y_0 which leads to a curve of the form B in Fig. 149 provides for a very close approximation to linear phase shift in the transmission range. In fact, Bode starts with this as the primary object and then shows that the resulting parameter determination at the same time provides for a y_0-curve of the form B in the attenuation range. In this case the at-

[1] *Loc. cit.*

tenuation function rises toward infinity uniformly, the rapidity of the rise beyond cut-off depending upon the number of factors in the index function. This number, of course, also determines the degree with which linearity is approximated by the phase function in the transmission range. A practical disadvantage of the Bode method lies in the fact that it does not utilize the function to its best advantage from the standpoint of tolerance and coverage so that a given rate of cut-off requires a more extensive network than on the Cauer basis.[1] Another practical difference between the two methods is the fact that the resulting lattice on the Cauer basis can always be developed into an equivalent ladder while this cannot be done completely when the parameter determination is carried out by Bode's method.

In the case of the oscillatory function, the degree of approximation is again most conveniently expressed in terms of a tolerance and coverage. Since unity is the *geometric mean* between the maxima and minima, the tolerance is given either as the ratio of the maximum positive deviation to unity, or the maximum negative deviation to the minimum of y_0. Thus denoting the tolerance by ϵ we have, according to Fig. 149,

$$\epsilon = y_0' - 1 = \frac{1 - y_0''}{y_0''}, \tag{861}$$

which is consistent with (857). The same applies to the function z_0. The tolerance can be defined also as the maximum *logarithmic* deviation, i.e., as $\ln(1 + \epsilon)$. For small tolerances this is very nearly the same as ϵ, so that it makes little difference, practically, which definition is used.

The minimum attenuation according to (858) and (861) is given in terms of the tolerance by

$$\gamma_{1\min} = \ln\left(\frac{2 + \epsilon}{\epsilon}\right), \tag{862}$$

or the tolerance in terms of the minimum attenuation by

$$\epsilon = \frac{2}{e^{\gamma_{1\min}} - 1}. \tag{862a}$$

4. Some useful lattice properties. Several interesting properties of the lattice structure may be seen by inspection of the relations (846). For example, if in a given structure we replace Z_a by R^2/Z_a *and* Z_b by $R^2/$

[1] It should be remembered, however, that where a linear phase characteristic is essential the Cauer filter would have to be followed by an all-pass phase corrective network. The combination of the two would then show a loss in economy as compared with the Bode filter, which accomplishes both ends simultaneously.

Z_b, the resulting lattice will have the propagation and characteristic impedance functions

$$\left.\begin{aligned} \gamma^* &= \ln\left(\frac{1+\sqrt{\dfrac{Z_b}{Z_a}}}{1-\sqrt{\dfrac{Z_b}{Z_a}}}\right) \\ \text{and} \qquad Z_0^* &= \frac{R^2}{\sqrt{Z_b Z_a}} = \frac{R^2}{Z_0} \end{aligned}\right\}. \tag{863}$$

If Z_a and Z_b are so determined that Z_0 approximates R in the Tscheby-scheff manner, then the reciprocal function Z_0^* approximates R in the same manner with the same tolerance and coverage. The propagation function γ^* differs from γ only in that its imaginary part is changed by $\pm\,\pi$. The attenuation for the filter is unchanged. The same is true for the type of approximation given by Bode's method of parameter determination. Hence replacing the reactances Z_a and Z_b by their reciprocals with respect to R^2 leads to a filter with the same properties regarding its practical usefulness. It should be noted also in this connection that *interchanging* Z_a and Z_b leaves Z_0 unchanged and replaces γ by γ^*. Hence the selective properties are not affected.

On the other hand, if Z_a be replaced by R^2/Z_a *or* Z_b by R^2/Z_b, which is the same as replacing z_a by $1/z_a$ *or* z_b by $1/z_b$, then either of the following pair of transformations takes place

$$\left.\begin{aligned} y_0 &\rightarrow \sqrt{z_b z_a} = z_0 \\ \text{and} \qquad z_0 &\rightarrow \sqrt{z_b/z_a} = y_0 \end{aligned}\right\},$$

or

$$\left.\begin{aligned} y_0 &\rightarrow 1/\sqrt{z_b z_a} = 1/z_0 \\ \text{and} \qquad z_0 &\rightarrow \sqrt{z_a/z_b} = 1/y_0 \end{aligned}\right\}.$$

That is, y_0 is replaced by z_0 *or* $1/z_0$, and z_0 is replaced by y_0 *or* $1/y_0$. In either case the filter is replaced by its *complement;* i.e., its transmission and attenuation ranges are interchanged. The complementary filter thus arrived at shows the same tolerance and coverage in its characteristic impedance as the original filter does for its index function, and vice versa. Thus a high-pass filter is obtained from a low-pass by reciprocating *either* Z_a or Z_b with respect to R^2, while the same procedure will change a band-pass into a band-elimination filter.

5. Impedance and index functions for various filter classes. In order to develop a design procedure for a filter of this kind, we consider separately the various classes such as low-pass, high-pass, etc., and develop a series of functions y_0 and z_0 for each whose complexity ranges

from the simplest form to that capable of meeting the most exact requirements which may be specified within the province of this method of approach. For each function in these series the objective is to draw up (wherever practicable) curves of tolerance and coverage versus parameter values by means of which the proper function and its parameters may be chosen readily to meet the requirements of a specific case. The function (855), for example, is the *fourth* in a series of y_0 for the low-pass class. The first in this series consists of the irrational factor $\sqrt{\omega_1^2 - \omega^2}$ alone; the second has in addition the factor $(\omega_a^2 - \omega^2)$ in the denominator, the third the factor $(\omega_b^2 - \omega^2)$ in the numerator, etc. The function (856) likewise is the fourth in a series of z_0 for the low-pass class. So long as these functions are so formed as to have simple poles and zeros, with the exception of the irrational factor, in those regions where they are imaginary, with alternate zeros and poles, the product or quotient of any z_0-function with any y_0-function of the same class will lead to physically realizable reactance functions z_b and z_a.

From what has been said above it is unnecessary to consider complementary classes separately. Thus, for example, a z_0-function for the low-pass class is at the same time a y_0-function for the high-pass class. Tables I and II give a series of functions for the **L.P.** and **H.P.** classes. The critical frequencies for the functions of both tables are subject to the condition

$$0 < \cdots \omega_b < \omega_a < \omega_1 < \omega_2 < \omega_3 \cdots < \infty, \tag{864}$$

but are otherwise arbitrary. H is a numerical factor whose significance

	TABLE I $\sqrt{z_b z_a}$ for **L.P.** or $\sqrt{z_b/z_a}$ for **H.P.**		TABLE II $\sqrt{z_b z_a}$ for **H.P.** or $\sqrt{z_b/z_a}$ for **L.P.**
2)	$\dfrac{H\sqrt{\omega_1^2 - \omega^2}}{\omega_1}$	2)	$\dfrac{H\sqrt{\omega_1^2 - \omega^2}}{j\omega}$
4)	$\dfrac{\omega_2^2\sqrt{\omega_1^2 - \omega^2}}{H\omega_1(\omega_2^2 - \omega^2)}$	4)	$\dfrac{j\omega\sqrt{\omega_1^2 - \omega^2}}{H(\omega_a^2 - \omega^2)}$
6)	$\dfrac{H\omega_2^2\sqrt{\omega_1^2 - \omega^2}(\omega_3^2 - \omega^2)}{\omega_1^2\omega_3^2(\omega_2^2 - \omega^2)}$	6)	$\dfrac{H\sqrt{\omega_1^2 - \omega^2}(\omega_b^2 - \omega^2)}{j\omega(\omega_a^2 - \omega^2)}$
8)	$\dfrac{\omega_2^2\omega_4^2\sqrt{\omega_1^2 - \omega^2}(\omega_3^2 - \omega^2)}{H\omega_1\omega_3^2(\omega_2^2 - \omega^2)(\omega_4^2 - \omega^2)}$	8)	$\dfrac{j\omega\sqrt{\omega_1^2 - \omega^2}(\omega_b^2 - \omega^2)}{H(\omega_a^2 - \omega^2)(\omega_c^2 - \omega^2)}$
. .	. . . .	.	. . . .

will become clear in the following discussion. ω_1 is the cut-off frequency. The student should investigate these functions to see that they are real

in their respective approximation ranges and otherwise behave like physical reactances.

The tables can be continued easily by adding functions with increasing complexity. The formation of these should be clear from inspection of those given. For the design of a low-pass filter any function may be selected for z_0 from Table I and combined with any y_0-function selected from Table II.[1] For a high-pass filter z_0 is selected from Table II and y_0 from Table I.

Two series of functions from which to select for the **B.P.** and **B.E.** classes are given in Tables III and IV. For the design of a band-pass

	TABLE III $\sqrt{z_b z_a}$ for **B.P.** or $\sqrt{z_b/z_a}$ for **B.E.**		TABLE IV $\sqrt{z_b z_a}$ for **B.E.** or $\sqrt{z_b/z_a}$ for **B.P.**
1)	$\dfrac{j\omega\sqrt{\omega_1^2-\omega^2}}{\omega_0\sqrt{\omega_{-1}^2-\omega^2}}$	1)	$\sqrt{\dfrac{\omega_1^2-\omega^2}{\omega_{-1}^2-\omega^2}}$
2)	$\dfrac{H\sqrt{(\omega_{-1}^2-\omega^2)(\omega_1^2-\omega^2)}}{j\omega w_1}$	2)	$\dfrac{H\sqrt{(\omega_{-1}^2-\omega^2)(\omega_1^2-\omega^2)}}{(\omega_0^2-\omega^2)}$
3)	$\dfrac{\omega_0(\omega_{-2}^2-\omega^2)\sqrt{\omega_1^2-\omega^2}}{j\omega\sqrt{\omega_{-1}^2-\omega^2}(\omega_2^2-\omega^2)}$	3)	$\dfrac{(\omega_{-a}^2-\omega^2)\sqrt{\omega_1^2-\omega^2}}{\sqrt{\omega_{-1}^2-\omega^2}(\omega_a^2-\omega^2)}$
4)	$\dfrac{j\omega w_2^2\sqrt{(\omega_{-1}^2-\omega^2)(\omega_1^2-\omega^2)}}{Hw_1(\omega_{-2}^2-\omega^2)(\omega_2^2-\omega^2)}$	4)	$\dfrac{\sqrt{\omega_{-1}^2-\omega^2}(\omega_0^2-\omega^2)\sqrt{\omega_1^2-\omega^2}}{H(\omega_{-a}^2-\omega^2)(\omega_a^2-\omega^2)}$
5)	$\dfrac{j\omega(\omega_{-2}^2-\omega^2)\sqrt{\omega_1^2-\omega^2}(\omega_3^2-\omega^2)}{\omega_0(\omega_{-3}^2-\omega^2)\sqrt{\omega_{-1}^2-\omega^2}(\omega_2^2-\omega^2)}$	5)	$\dfrac{(\omega_{-a}^2-\omega^2)(\omega_b^2-\omega^2)\sqrt{\omega_1^2-\omega^2}}{\sqrt{\omega_{-1}^2-\omega^2}(\omega_{-b}^2-\omega^2)(\omega_a^2-\omega^2)}$
6)	$\dfrac{Hw_2^2(\omega_{-3}^2-\omega^2)\sqrt{(\omega_{-1}^2-\omega^2)(\omega_1^2-\omega^2)}(\omega_3^2-\omega^2)}{j\omega w_1 w_3^2(\omega_{-2}^2-\omega^2)(\omega_2^2-\omega^2)}$	6)	$\dfrac{H\sqrt{\omega_{-1}^2-\omega^2}(\omega_{-b}^2-\omega^2)(\omega_b^2-\omega^2)\sqrt{\omega_1^2-\omega^2}}{(\omega_{-a}^2-\omega^2)(\omega_0^2-\omega^2)(\omega_a^2-\omega^2)}$
. . .	. . . .	. . .	. . . .

filter the z_0-function is selected from Table III and the y_0-function from Table IV. For the band-elimination filter Table IV is used for z_0 and Table III for y_0. For both tables the critical frequencies separate each other according to the relation

$$0 < \cdots \omega_{-3} < \omega_{-2} < \omega_{-1} < \omega_{-a} < \omega_{-b} \cdots < \omega_0 < \cdots$$
$$\omega_b < \omega_a < \omega_1 < \omega_2 < \omega_3 \cdots < \infty, \qquad (865)$$

[1] As will be seen in the following discussion, the factor H in general has different values for the y_0 and z_0 functions selected from the two tables and hence should be supplied with suitable distinguishing subscripts, for example, as H_y and H_z. For obvious reasons such a distinction is not practicable in the above tables.

but are otherwise arbitrary. ω_{-1} and ω_1 are, respectively, the left- and right-hand boundaries of the pass or attenuation bands. H is again a numerical factor, and w_1, w_2, etc., are functions of the critical frequencies. Their significance will be discussed in the next section. Since they are constant they have no effect upon the general character of the functions. The manner in which the series in either table may be continued is clarified somewhat by reference to Fig. 150. Part (a) of this figure shows the general distribution of poles and zeros for the odd-numbered functions. For these the boundary frequencies ω_{-1} and ω_1 correspond to a pole and zero, respectively, owing to the appearance of an irrational factor in the denominators and numerators of such functions. The irrational character is indicated by the figures $\frac{1}{2}$ placed above these critical frequencies. In part (b) the situation is similarly represented for the even-numbered functions for which both boundaries correspond to zeros. The series in Table III is continued by forming

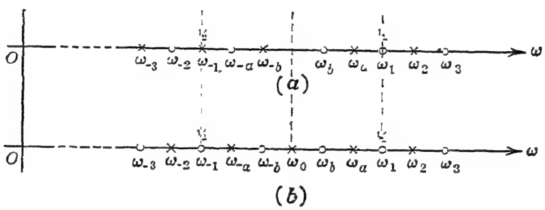


FIG. 150.—Distribution of critical frequencies for the odd (a) and even (b) functions of Tables III and IV.

functions for which additional pairs of critical frequencies are added outside the interval ω_{-1} to ω_1, while for the continuation of Table IV additional pairs are introduced within this interval.

It is significant here to point out that the even-numbered functions in Tables III and IV possess the same ability to approximate unity in their respective approximation ranges as the correspondingly numbered functions in Tables I and II. We shall see later that the even-numbered functions in Tables III and IV do for the **B.P.** and **B.E.** classes what the similarly numbered functions in Tables I and II do for the **L.P.** and **H.P.** classes. The odd-numbered functions in Tables III and IV possess approximation abilities which place them in the positions given. They exist only for the **B.P.** and **B.E.** classes and hence cannot be formed for inclusion in Tables I and II. This may be appreciated in a general way from the fact that the boundaries in the odd-numbered functions are formed by a pole and zero, respectively, while for the even-numbered functions these are both zeros. Since the **L.P.** and **H.P.** filters have only one cut-off point in the positive frequency axis, the possibility of form-

ing analogous functions to the odd-numbered ones does not exist. Incidentally, the process of replacing poles by zeros and vice versa in all these functions merely leads to the characteristic impedance and propagation functions Z_0^* and γ^*, and hence presents no alternative possibilities. As far as the tables are carried, therefore, they present a complete list from which to select.

6. Frequency transformations. The study of the properties of the above functions is greatly simplified by introducing suitable frequency transformations. By means of such changes of the independent variable, the functions for the four filter classes appearing in the above tables are reduced to a single set. A study of these then suffices for the parameter determination in the design of any of the four common classes of filters.[1] The transformations applicable to the functions in Tables I to IV are summarized in Table V. The symbol x is used to designate the new

TABLE V

			Odd Functions		Even Functions	
	Table I	Table II	Table III	Table IV	Table III	Table IV
x	$\frac{\omega}{\omega_1}$	$\frac{\omega_1}{\omega}$	$\frac{\omega^2 - \omega_0^2}{w_1\omega_0}$	$\frac{w_1\omega_0}{\omega^2 - \omega_0^2}$	$\frac{\omega^2 - \omega_0^2}{w_1\omega}$	$\frac{w_1\omega}{\omega^2 - \omega_0^2}$
x_ν	$\frac{\omega_\nu}{\omega_1}$	$\frac{\omega_1}{\omega_\nu}$	$\frac{w_\nu}{w_1} = \frac{\omega_\nu^2 - \omega_0^2}{w_1\omega_0}$	$\frac{w_1}{w_\nu} = \frac{w_1\omega_0}{\omega_\nu^2 - \omega_0^2}$	$\frac{w_\nu}{w_1} = \frac{\omega_\nu^2 - \omega_0^2}{w_1\omega_\nu}$	$\frac{w_1}{w_\nu} = \frac{w_1\omega_\nu}{\omega_\nu^2 - \omega_0^2}$
w_ν	. . . . .		$\frac{\omega_\nu^2 - \omega_{-\nu}^2}{2\omega_0}$		$\omega_\nu - \omega_{-\nu}$	
ω_0	. . . . .		$\sqrt{\frac{\omega_\nu^2 + \omega_{-\nu}^2}{2}}$		$\sqrt{\omega_\nu\omega_{-\nu}}$	

For all classes $\nu = 1, 2, 3, \ldots$ or $a, b, c, \ldots$

variable. w_ν and ω_0 (appearing also in Tables III and IV) are parameters by means of which the substitutions take on more convenient forms. The terms "odd" and "even" refer to the odd- and even-numbered functions. For the latter, w_ν is the "width" of the angular frequency in-

[1] This applies particularly to such design problems for which the Cauer type of parameter determination is suitable. Where the Bode type of parameter determination is called for on account of a linear phase requirement, the problem involves more detailed considerations which will be pointed out later.

terval between a pair of corresponding critical frequencies, while ω_0 is their geometric mean. For the odd functions ω_0 is the root-mean-square between such frequency pairs, and w_ν is the width multiplied by a factor which is the ratio of the arithmetic mean to the root-mean-square; i.e.,

$$w_\nu = (\omega_\nu - \omega_{-\nu}) \cdot \left(\frac{\omega_\nu + \omega_{-\nu}}{2\omega_0}\right).$$

As the ratio of band width to mean frequency decreases, these relations become practically the same as those for the even functions, which were also used in connection with the conventional filter theory discussed in the previous chapter. The variable x for the even functions is practically the same as x_k which was used in the discussion of the constant-k type. The parameters x_ν are those appearing in the transformed functions, the first six of which are given in Table VI. These are numbered to correspond to the similarly numbered functions in Tables I to IV. Thus any of the functions in Table VI may be identified with corresponding

TABLE VI *

$F_1(x) = \sqrt{\dfrac{1-x}{1+x}}$	$F_2(x) = H\sqrt{1-x^2}$
$F_3(x) = \sqrt{\dfrac{1-x}{1+x}} \cdot \left(\dfrac{x_2+x}{x_2-x}\right)$	$F_4(x) = \dfrac{x_2^2\sqrt{1-x^2}}{H(x_2^2-x^2)}$
$F_5(x) = \sqrt{\dfrac{1-x}{1+x}} \cdot \left(\dfrac{x_2+x}{x_2-x}\right) \cdot \left(\dfrac{x_3-x}{x_3+x}\right)$	$F_6(x) = \dfrac{Hx_2^2\sqrt{1-x^2}(x_3^2-x^2)}{x_3^2(x_2^2-x^2)}$
. . . .	. . . .

* NOTE: For Tables II and IV replace x_2, x_3, etc., by x_a, x_b, etc.

functions in the previous tables by applying the proper transformation as given in Table V. For example, suppose we wish to identify $F_4(x)$ with function 4) of Table III. The change of variable is given as

$$x = \frac{\omega^2 - \omega_0^2}{w_1\omega} = \frac{\omega^2 - \omega_0^2}{\omega(\omega_1 - \omega_{-1})}.$$

Hence

$$1 - x^2 = \frac{\omega^2(\omega_1 - \omega_{-1})^2 - \omega^4 - \omega_0^4 + 2\omega_0^2\omega^2}{w_1^2\omega^2}.$$

But since $\omega_0^2 = \omega_1\,\omega_{-1}$, this is

$$1 - x^2 = \frac{\omega^2(\omega_1^2 + \omega_{-1}^2) - \omega^4 - \omega_0^4}{w_1^2\omega^2}.$$

Similarly, noting that $x_2 = w_2/w_1$, we have

$$x_2^2 - x^2 = \frac{(\omega_{-2}^2 - \omega^2)(\omega_2^2 - \omega^2)}{- w_1\omega^2}.$$

Substitution of these into $F_4(x)$ leads to function 4) of Table III as may easily be verified.

The above transformations are so chosen that the approximation range corresponds to $-1 \leq x \leq 1$ in every case. This is the range in which the functions $F_1(x)$ to $F_6(x)$ are real and positive. The reduction to the functions of Table VI is exact in all cases except for functions 1, 3, 5, etc., of Table III, which become

$$\left(\frac{\omega}{\omega_0}\right)F_1(x); \left(\frac{\omega_0}{\omega}\right)F_3(x); \left(\frac{\omega}{\omega_0}\right)F_5(x) \quad \cdots .$$

If the band width is small compared with the mid-band frequency ω_0, the factor ω/ω_0, which varies from ω_{-1}/ω_0 to ω_1/ω_0 as x varies from -1 to 1, is very nearly unity throughout the approximation range. When this condition does not hold, the parameter determination by means of the functions $F_1(x)$, $F_3(x)$, . . . may be used as a first approximation in a subsequent cut-and-try process. Incidentally, it is interesting to note that the odd-numbered functions are odd functions of x, and the even-numbered ones are even.

7. Cauer's method of parameter determination. By this method the parameters x_2, x_3, etc., in the functions of Table VI are so determined that the behavior of the latter becomes oscillatory in the general manner illustrated by curve A of Fig. 149.[1] The line of reasoning to be given here rests upon the suggestion that this oscillatory form of behavior resembles that of a periodic function with the exception that the periods become shortened as the cut-off points are approached, and that the quasi-periodic character is limited to the interval $-k < x < k$. It seems reasonable that both of these exceptions may be overcome by a suitable change of variable. This is illustrated in Fig. 151 in which the variable x is plotted horizontally and a new variable u vertically. The change of variable expressed by $x = f(u)$ is a periodic function which oscillates between the values $\pm k$. As u varies continuously, therefore, x retraces the region $-k$ to k. At the same time the form of the function $f(u)$ is such that, when $F(x)$ is projected through it, the resulting $F(u)$ becomes periodic. This process of projection is indicated in the figure

[1] In his publication, "Siebschaltungen," Cauer states this manner of parameter determination without proof. The discussion given here is not intended to be a rigorous demonstration but rather a presentation of the general idea involved. It is a modification of a rigorous derivation due to Mr. S. A. Schelkunoff made available to the writer through the courtesy of H. W. Bode of the Bell Telephone Laboratories.

by the dotted lines. The amplitude of oscillation of $F(u)$ is chosen arbitrarily and has nothing to do with the idea which the figure is intended to convey.

If the transformation $x = f(u)$ can be found, then the process of determining the parameters $x_2, x_3, \ldots$ in the functions $F(x)$ of Table VI becomes one of making $F(u)$ periodic. This we shall now discuss in

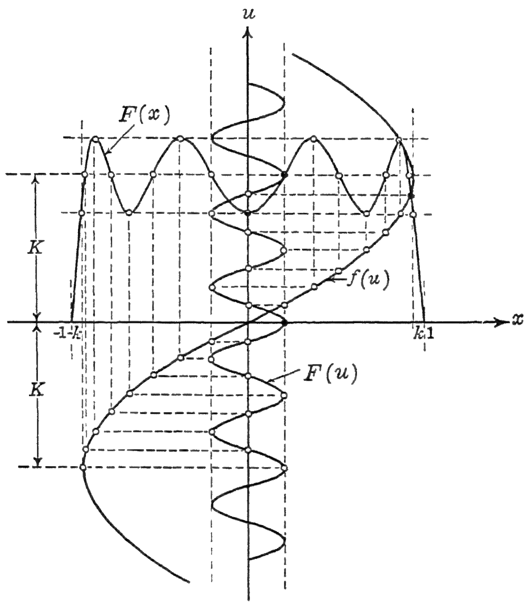


FIG. 151.—Change of variable $x = f(u)$ which transforms the functions $F(x)$ of Table VI into periodic functions $F(u)$ over the range $-k \leq x \leq k$.

more detail for the even functions. These have the following general form

$$F_{2n}(x) = H^{\pm 1} \cdot \frac{\sqrt{1 - \frac{x^2}{x_1^2}}\left(1 - \frac{x^2}{x_3^2}\right) \cdots \left(1 - \frac{x^2}{x_n^2}\right)}{\left(1 - \frac{x^2}{x_2^2}\right)\left(1 - \frac{x^2}{x_4^2}\right) \cdots \left(1 - \frac{x^2}{x_{n-1}^2}\right)}, \tag{866}$$

$$(x_1 < x_2 < x_3 \cdots x_n < \infty)$$

in which $x_1 = 1$ is introduced for the sake of uniformity, and the sign in the exponent of H is plus or minus according to whether n is *odd* or *even*.

Since the desired transformation must be a periodic function which

oscillates between $\pm k$, it seems natural to try tentatively the trigonometric sine and write

$$x = k \sin \varphi. \tag{867}$$

If the parameters are written as[1]

$$x_\nu = \frac{1}{\sin \varphi_\nu}, \tag{868}$$

then the factors in (866) have the form

$$\left(1 - \frac{x^2}{x_\nu^2}\right) = (1 + k \sin \varphi \sin \varphi_\nu)(1 - k \sin \varphi \sin \varphi_\nu). \tag{869}$$

Let us suppose for the moment that the transformation (867) is the correct one, i.e., that φ may be considered the same as the desired variable u. If the parameters φ_ν, as we pass from one factor in the numerator or denominator of (866) to the next, follow a sequence with equal intervals such that the factors (after some rearrangement perhaps) form a closed cycle, i.e., form such a group that the one logically following the last becomes identical to the first, then the same intervals (increments) in φ leave the set of factors as a whole unaffected provided the individual factors are unchanged by equal increments in φ and φ_ν in the same or opposite directions. Such increments in φ merely amount to shifting the positions of the various factors to the right or left around the cycle. They thus leave the function F unchanged; i.e., F is periodic and the equal increments are its periods.

Although the actual situation is somewhat different, the above line of thought has in it the essentials involved in the process of making $F(u)$ periodic and in finding the proper transformation function $f(u)$. Having the factors remain unchanged by equal increments in φ and φ_ν is a sufficient condition for bringing about the periodicity of F but not necessary, and hence perhaps too severe. The function F is determined (except for a constant multiplier) by the zeros of its numerator and denominator. Hence it is only necessary that the location of a zero[2] defined by any factor remain unchanged by equal increments in φ and φ_ν. Let us, therefore, see in what manner φ and φ_ν may vary and leave such a zero unchanged. Mathematically we wish to determine the condition on φ and φ_ν demanded by

$$\left.\begin{array}{lr} & d(1 \pm k \sin \varphi \sin \varphi_\nu) = 0 \\ \text{for} & 1 \pm k \sin \varphi \sin \varphi_\nu = 0 \end{array}\right\}. \tag{870}$$

[1] Since the parameters are all larger than unity, the reciprocal form is necessary.

[2] The zeros lie outside the approximation range, of course, and do not occur, therefore, for real values of φ.

Forming the differential and equating it to zero gives

$$k \cos \varphi \sin \varphi_\nu \, d\varphi + k \sin \varphi \cos \varphi_\nu \, d\varphi_\nu = 0. \tag{871}$$

But by the second relation (870)

$$\cos \varphi = \frac{\pm j\sqrt{1 - k^2 \sin^2 \varphi_\nu}}{k \sin \varphi_\nu}; \quad \cos \varphi_\nu = \frac{\pm j\sqrt{1 - k^2 \sin^2 \varphi}}{k \sin \varphi}, \tag{872}$$

in which the + and − signs do not have to be taken in pairs. Substitution into (871) gives the desired condition

$$\frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}} \pm \frac{d\varphi_\nu}{\sqrt{1 - k^2 \sin^2 \varphi_\nu}} = 0. \tag{873}$$

Although this result is contrary to what was supposed in the above reasoning, it does indicate a further change of variable which will bring about the desired end. What we need in place of (873) is a condition of the form

$$du \pm du_\nu = 0, \tag{873a}$$

which makes it evident that the proper relation between φ and u is given by

$$du = \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}}. \tag{874}$$

Since this indicates that $du/d\varphi$ is always positive, we see that u is a uniformly increasing function of φ. If we specify that $u = 0$ for $\varphi = 0$, then the desired functional relationship is given by the integral

$$u = \int_0^\varphi \frac{d\xi}{\sqrt{1 - k^2 \sin^2 \xi}}, \tag{875}$$

which is known as an *elliptic integral* of the first kind. u as a function of φ depends upon the parameter k. Its general form is given in Fig. 152. At φ equal to integer multiples of π (inclusive of zero) the slope is unity, while at odd multiples of $\pi/2$ it has its maximum value $(1 - k^2)^{-\frac{1}{2}}$. The value K corresponds to $\varphi = \pi/2$ and is called the *complete elliptic integral* of the first kind. It varies from $\pi/2$ to ∞ as k varies from zero to unity.

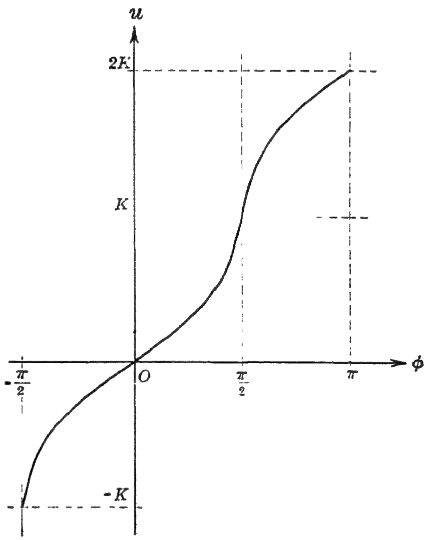


FIG. 152.—Plot showing the general character of the elliptic integral (875) as a function of its upper limit.

Evidently we may also consider (875) as defining φ as a function of u. The student will perhaps appreciate this more easily by reference to Fig. 152. Thus with (867) the required transformation $x = f(u)$ is given. Mathematically this is written as

$$x = k \sin \varphi = k\, sn\, u, \tag{867a}$$

in which the *sn-function* is the Jacobian elliptic function. According to the above discussion it has a period equal to $4K$ which depends upon the parameter k. For $k = 0$, $u = \varphi$, and the *sn*-function is identical to the trigonometric sine; while for $k = 1$ its period is infinite. For an inter-

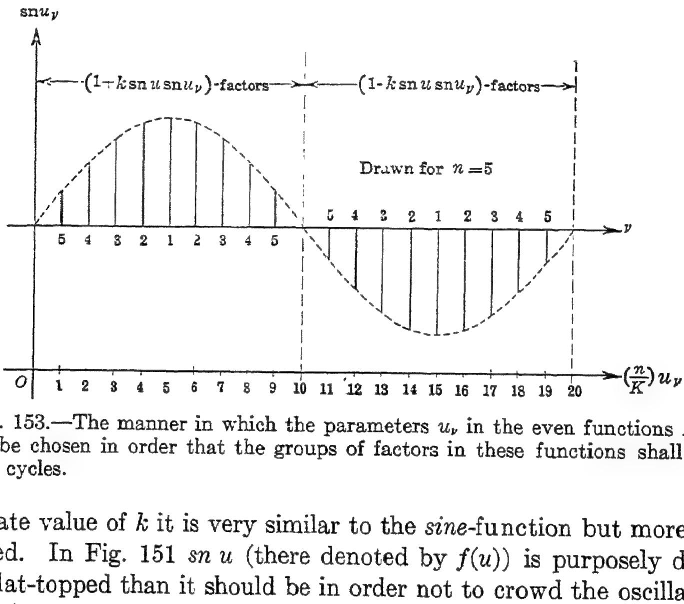


FIG. 153.—The manner in which the parameters u_ν in the even functions $F_{2n}(u)$ must be chosen in order that the groups of factors in these functions shall form closed cycles.

mediate value of k it is very similar to the *sine*-function but more flat-topped. In Fig. 151 $sn\, u$ (there denoted by $f(u)$) is purposely drawn less flat-topped than it should be in order not to crowd the oscillations of $F(x)$ too closely together at the cut-off.

The parameters are now expressed with the help of (868) by

$$x_\nu = \frac{1}{sn\, u_\nu}, \tag{868a}$$

in which the u_ν are now to be chosen so as to establish the cyclic character in the groups of factors in $F(x)$ described earlier. These factors according to (869) are now written

$$\left(1 - \frac{x^2}{x_\nu^{\,2}}\right) = (1 + k\, sn\, u\, sn\, u_\nu)(1 - k\, sn\, u\, sn\, u_\nu). \tag{869a}$$

It is easy to see that $F_{2n}(x)$ has n complete periods in the interval $-k < x < k$. The transformed function $F_{2n}(u)$, therefore, has n periods

in the interval $-K < u < K$, so that a period is equal to $2K/n$ in terms of u. This, then, must be the uniform increment in the u_ν for the factors of the numerator or denominator.

The manner in which the u_ν-values must be chosen in order that the groups of factors in the numerator and denominator of F form closed cycles is best seen by reference to Fig. 153 which is a plot of $sn\, u_\nu$ versus ν or $\left(\frac{n}{K}\right)u_\nu$ for the case $n = 5$. The vertical lines represent the magnitudes of $sn\, u_\nu$. It is clear from (869a) that each group of factors splits into two groups, one involving factors with positive signs and the other with negative signs. These we shall refer to as the positive and negative groups, respectively, and consider the positive first. Since $1/sn\, u_1 = x_1$ is contained in the irrational factor of (866) and, therefore, must equal unity, u_1 must equal K; i.e., a quarter period of the sn-function. The remaining u_ν-values in the positive group must then be so chosen as to divide this quarter period into n equal parts. The positive maximum ordinate of the $sn\, u_\nu$-curve, therefore, corresponds to $\nu = 1$, and the remaining ordinates of the first quarter-period at equal intervals correspond to $\nu = 2, 3, 4, 5$ from right to left.

This, however, does not check with the fact that the factor for $\nu = 1$ is contained under a radical while the others are not. We can overcome this difficulty by squaring the function $F_{2n}(u)$. Then the factor with $\nu = 1$ occurs once and the remaining ones twice. The latter are easily represented by the ordinates $\nu = 2, 3, 4, 5$ in the second quarter-period, with corresponding ones in the first. The total of the ordinates in the first half-period then represent all the factors in the positive groups in the numerator and denominator, it being understood that alternate ordinates correspond to factors in the numerator and denominator groups so that the increments in u_ν for either group become $2K/n$ as required. In exactly the same manner the $sn\, u_\nu$ in the factors of the negative groups are given by ordinates in the second half-period. Since these are negative, the corresponding factors may be written with plus signs provided the arguments u_ν are expressed consecutively as indicated in the lower axis of Fig. 153.

It thus becomes clear that, *for n-odd*, we have

$$F_{2n}^2(u) = H^2 \frac{\left(1 + k\, sn\, u\, sn\frac{K}{n}\right)\left(1 + k\, sn\, u\, sn\frac{3K}{n}\right) \cdots \left(1 + k\, sn\, u\, sn\frac{4n-1}{n}K\right)}{\left(1 + k\, sn\, u\, sn\frac{2K}{n}\right)\left(1 + k\, sn\, u\, sn\frac{4K}{n}\right) \cdots \left(1 + k\, sn\, u\, sn\frac{4n}{n}K\right)} \tag{876}$$

while the parameters u_ν (of which there are n) are given by

$$u_\nu = \left(\frac{n - \nu + 1}{n}\right)K; \; (\nu = 1, 2, 3, \cdots n). \tag{877}$$

This may be seen to check with the order of factors illustrated in Fig. 153 which represents a case for n-odd. When n is *even* the same line of reasoning leads to the reciprocal of (876).[1] The student may carry this case through as an exercise.

In order to recognize the periodicity of (876), suppose we increase u by $2K/n$. Since the locations of the zeros defined by the factors remain unchanged by a subsequent decrease in u by $2K/n$ and a simultaneous decrease or increase in u_ν by the same amount, such a subsequent change in u and u_ν will merely shift the order of the factors around the cycle in the right- or left-hand directions, respectively. Note, incidentally, that the last factor in the denominator of (876) equals unity and is added merely to complete the cycle.

By means of (868a) and (877) the parameters in $F_{2n}(x)$ are now given. They are

$$x_\nu = \frac{1}{sn\left(\frac{n - \nu + 1}{n}\right)K}; \; (\nu = 1, 2, \cdots n). \tag{878}$$

This parameter determination makes all the maxima and all the minima in $F_{2n}(x)$ equal. For n-odd the origin is a maximum, for n-even a minimum. The maxima and minima lie at half-period intervals and hence occur at

$$x = k\, sn \frac{\nu K}{n}; \; (\nu = 0, 1, 2, \cdots n), \tag{879}$$

the last of these being $x = k$ which is the point at which the function passes through the value equal to its minima. These are the points of minimum attenuation when $F_{2n}(x)$ represents an index function. k is the coverage. $H^{\pm 1}$ is the value of the function at the origin. Since the nominal value is taken as the geometric mean between the maxima and minima and is equal to unity, we have

$$H = F_{\max} = 1/F_{\min}, \tag{880}$$

while the tolerance is given by

$$\epsilon = H - 1. \tag{881}$$

[1] This shows that the value unity is the geometric mean between the maxima and minima, i.e., that the function approximates unity in the Tschebyscheff manner as described above.

The points at which the function equals unity are the odd quarter-periods[1] and hence correspond to

$$x = k\, sn \frac{\nu K}{2n}; \quad (\nu = 1, 3, 5, \cdots 2n - 1). \tag{882}$$

These are the locations of the points of infinite attenuation when $F_{2n}(x)$ represents an index function.

The parameter determination for the odd functions of Table VI is carried out in an analogous fashion. Their general form is

$$F_{2n-1}(x) = \frac{\sqrt{1 - \frac{x}{x_1}}\left(1 + \frac{x}{x_2}\right)\left(1 - \frac{x}{x_3}\right) \cdots \left(1 \pm \frac{x}{x_n}\right)}{\sqrt{1 + \frac{x}{x_1}}\left(1 - \frac{x}{x_2}\right)\left(1 + \frac{x}{x_3}\right) \cdots \left(1 \mp \frac{x}{x_n}\right)}, \tag{883}$$

in which the top signs in the last factors are for *n-even* and the bottom ones for *n-odd*. Here the function has $n - \frac{1}{2}$ or $\frac{2n-1}{2}$ complete periods in the interval $-k < x < k$. One period in terms of the variable u is, therefore, $\frac{4K}{2n-1}$. This is the uniform increment in u_ν for the factors of the numerator or denominator. Again recognizing that u_1 must equal K, and laying off increments of $\frac{2K}{2n-1}$ to the right and left of this point in a fashion similar to that illustrated in Fig. 153, we find for *n-odd or even*[2]

$$F_{2n-1}{}^2(u) = \frac{\left(1 + k\, sn\, u\, sn \frac{3K}{2n-1}\right)\left(1 + k\, sn\, u\, sn \frac{7K}{2n-1}\right) \cdots}{\left(1 + k\, sn\, u\, sn \frac{K}{2n-1}\right)\left(1 + k\, sn\, u\, sn \frac{5K}{2n-1}\right) \cdots} \cdot \frac{\left(1 + k\, sn\, u\, sn \frac{8n-5}{2n-1}K\right)}{\left(1 + k\, sn\, u\, sn \frac{8n-7}{2n-1}K\right)}, \tag{884}$$

and for the parameters

$$x_\nu = \frac{1}{sn\left(\frac{2(n-\nu)+1}{2n-1}\right)K}; \quad (\nu = 1, 2, 3, \cdots n), \tag{885}$$

which is the same as (878) with n replaced by $n - \frac{1}{2}$.

[1] This is exact only when the functions are so normalized that unity becomes the geometric mean between the maxima and minima as has been done here.

[2] In this case each quarter-period of $sn\ u_\nu$ is divided into $n-\frac{1}{2}$ equal increments.

The maxima and minima occur at

$$x = k\,sn\frac{\nu K}{2n-1}; \quad (\nu = 1, 3, 5, \cdots 2n-1), \tag{886}$$

and are the points at which minimum attenuation occurs[1] when $F_{2n-1}(x)$ represents an index function. At half-period intervals from the origin (inclusive of the latter) the function equals unity. These correspond to points of infinite attenuation. They are given by

$$x = k\,sn\frac{2\nu K}{2n-1}; \quad (\nu = 0, 1, 2, \cdots n-1). \tag{887}$$

k is again the coverage. The tolerance is expressed in the same manner as for the even functions. If we define H by means of (880), then the tolerance is given by (881). It is possible to express H for the various functions directly in terms of k and the parameters x_ν. This is done in Table VII.[2] By means of these relations the tolerance is more readily calculated for the higher order functions.

TABLE VII

Functions	$\frac{H^2-1}{H^2+1}$	$\sqrt{1-H^{-4}}$
$F_1(x)$	k	—
$F_2(x)$	—	k
$F_3(x)$	$k^3x_2^{-4}$	—
$F_4(x)$	—	$k^2x_2^{-4}$
$F_5(x)$	$k^5x_2^{-4}x_3^{-4}$	—
$F_6(x)$	—	$k^3x_3^{-4}$
$F_8(x)$	—	$k^4x_2^{-4}x_4^{-4}$
$F_{10}(x)$	—	$k^5x_3^{-4}x_5^{-4}$
$F_{12}(x)$	—	$k^6x_2^{-4}x_4^{-4}x_6^{-4}$

The process of numerical computation of the parameters from the relations (878) and (885) is accomplished by means of a table of elliptic integrals such as given by Jahnke and Emde,[3] for example. It is

[1] Unity is again the geometric mean between the maxima and minima, i.e., this method of parameter determination leads to the Tschebyscheff manner of approximation for both the odd and even functions.

[2] Taken from W. Cauer, "Siebschaltungen," p. 21.

[3] Jahnke-Emde: "Funktionentafeln mit Formeln und Kurven," B. G. Teubner, Berlin, 1928, pp. 53–59. These are four-place tables. A five-place table of sn-functions is found in "Fünfstellige Funktionentafeln," by Keiichi Hayashi, J. Springer, Berlin, 1930, pp. 128–135.

customary to put $k = \sin\alpha$ and write for the integral (875) $F(\alpha, \varphi)$. The tables give values of this function in terms of φ and the parameter

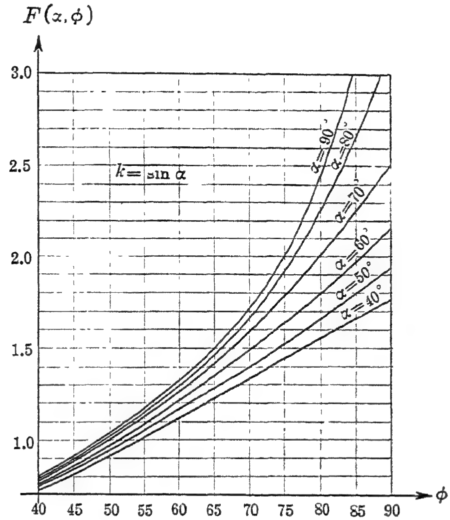


FIG. 154.—The elliptic integral (875) as a function of its upper limit for various values of the parameter k.

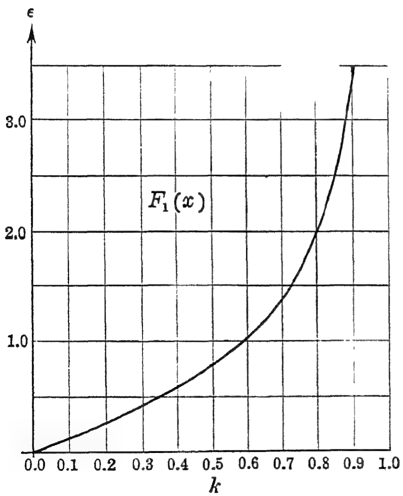


FIG. 155.—Combinations of tolerance (ϵ) and coverage (k) obtainable with $F_1(x)$ of Table VI.

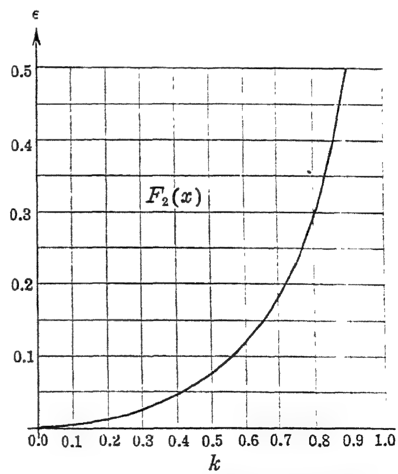


FIG. 156.—Combinations of tolerance (ϵ) and coverage (k) obtainable with $F_2(x)$ of Table VI.

α. In our case $u = F(\alpha, \varphi)$ and $\alpha = \sin^{-1}k$ are the given quantities, and the tables are used to find φ. With $sn\ u = \sin\varphi$, the parameters x_ν

are readily found. A family of curves representing $F(\alpha, \varphi)$ versus φ for various values of α are shown plotted in Fig. 154. Such a plot is easier to use than the tables, but far less accurate, of course.

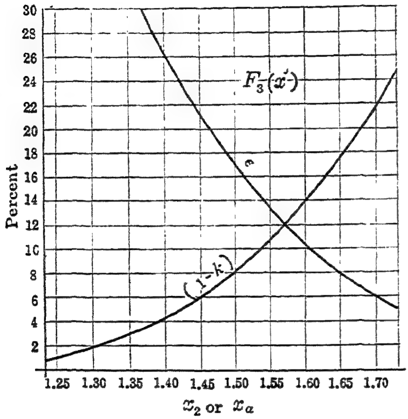


FIG. 157.—Tolerance and coverage obtainable for various values of the parameter in $F_3(x)$ of Table VI.

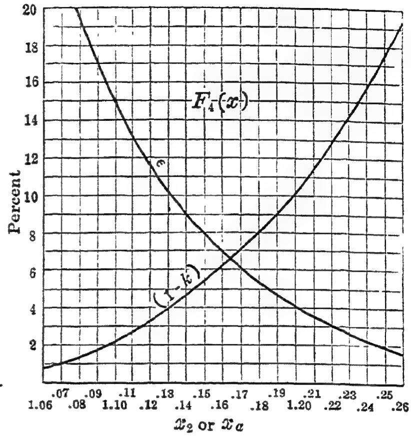


FIG. 158.—Tolerance and coverage obtainable for various values of the parameter in $F_4(x)$ of Table VI.

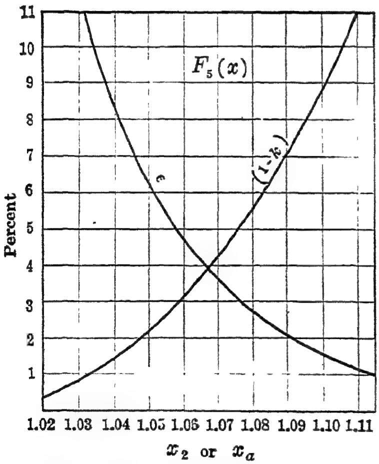


FIG. 159.—Combinations of tolerance and coverage obtainable with $F_5(x)$ of Table VI, and the corresponding values of the parameter x_2 or x_a.

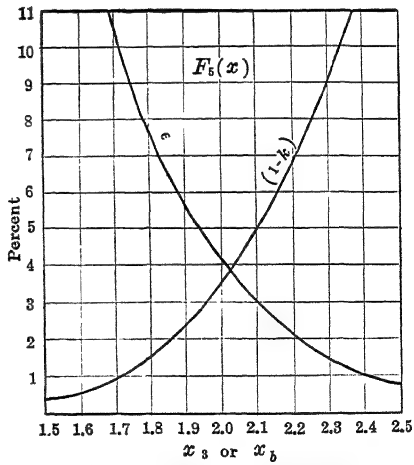


FIG. 160.—Curves corresponding to those of Fig. 159 for the determination of the parameter x_3 or x_b in $F_5(x)$.

Still more convenient for design purposes are plots of tolerance and coverage versus parameter values for the various functions. Figs. 155 to 162 show these relationships for functions $F_1(x)$ to $F_6(x)$ inclusive.

In all but the first two the percentage not covered $(1 - k)\cdot 100$ is plotted rather than the coverage $(k\cdot 100)$ in order that the same scale may be used for this quantity and the tolerance.

Fig. 163 shows plots of $F_1(x)$ to $F_6(x)$ which illustrate their comparative behaviors. The even functions are seen to be similar to W_{1k}, W_{1km}, $W_{1kmm'}$ . . . with which we are familiar from the design methods discussed in the previous chapter. The odd functions are found in the T- and Π-filter types discussed in section 13 of the previous chapter. The lattice structure includes these in its regular design procedure.

8. Illustrations. The following numerical examples may serve to clarify the method of filter design on the basis of Cauer's parameter

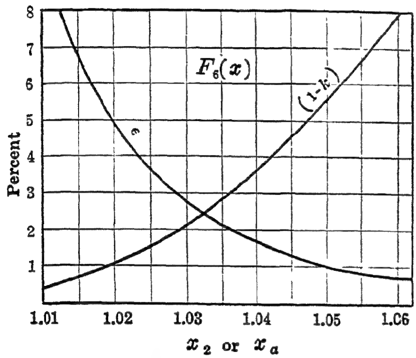


Fig. 161.—Combinations of tolerance and coverage obtainable with $F_6(x)$ of Table VI, and the corresponding values of the parameter x_2 or x_a.

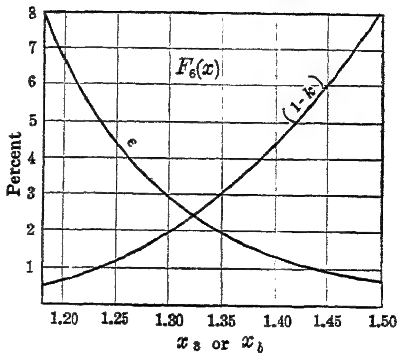


Fig. 162.—Curves corresponding to those of Fig. 161 for the determination of the parameter x_3 or x_b in $F_6(x)$.

determination. Suppose we wish to design a low-pass filter. To begin with, we note that the normalized characteristic impedance functions may be chosen from Table I, and the index functions from Table II. Turning to Table V, the frequency transformations which lead to the corresponding normalized functions of Table VI, therefore, are given by

$$\left.\begin{aligned} x &= \frac{\omega}{\omega_1} \\ x_\nu &= \frac{\omega_\nu}{\omega_1} \end{aligned}\right\} \tag{888}$$

for the characteristic impedance, and by

$$\left.\begin{aligned} x &= \frac{\omega_1}{\omega} \\ x_\nu &= \frac{\omega_1}{\omega_\nu} \end{aligned}\right\} \tag{888a}$$

for the index functions. Here ω_1 is the cut-off frequency. Let us suppose that the minimum attenuation is required to be 40 db, which corresponds to a numerical ratio of input to output voltage of 100 to 1. According to (862a) this fixes the tolerance at 2 per cent (taking into account the conversion from decibels to napiers). The point at which this minimum value of attenuation will first be reached depends upon

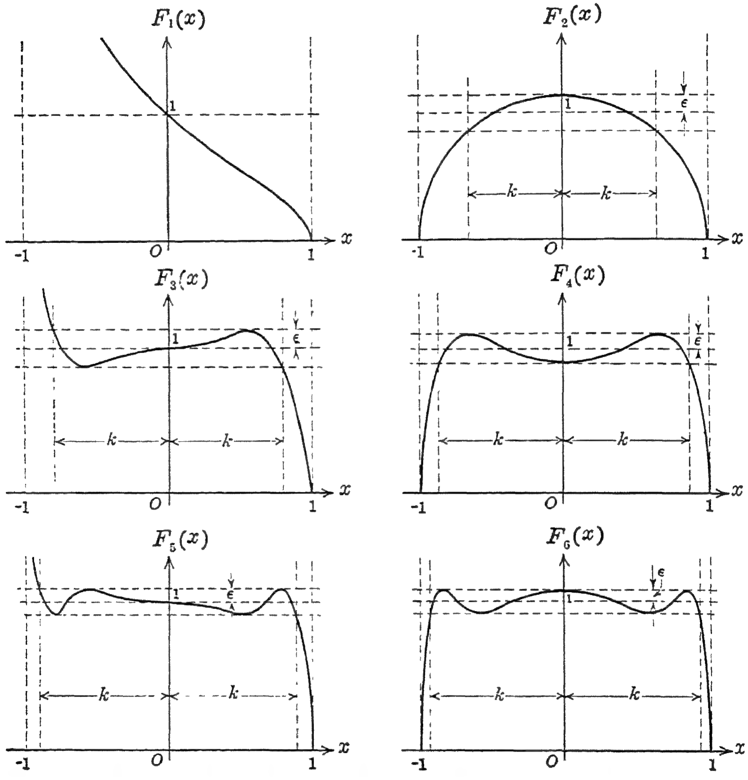


FIG. 163.—Curves showing the general character of the behavior of the functions $F_1(x)$ to $F_6(x)$ of Table VI when their parameters (with regard to tolerance ϵ and coverage k) are determined on the Tschebyscheff basis according to Figs. 155 to 162.

the corresponding coverage; namely, it is that frequency at which $x = k$, or using the first relation (888a) this is $\omega = \omega_1/k$. Thus if we consider $F_4(x)$ as a possible choice, Fig. 158 shows that $\epsilon = 0.02$ is obtained with $k = 0.825$ for the parameter value $x_a = 1.25$. For this choice the minimum attenuation will be reached, therefore, at $\omega = 1.212\omega_1$. The next better possibility is offered by $F_6(x)$, for which Fig. 162 shows that the coverage corresponding to $\epsilon = 0.02$ is expressed by

$k = 0.97$, so that the minimum attenuation with this function would be reached at $\omega = 1.031\omega_1$, which is an extremely sharp cut-off rate. Let us say that our design calls for a rate of cut-off which is sufficiently met by $F_4(x)$, i.e., that the sharpness of cut-off afforded by $F_6(x)$ is far more than required. We thus choose $F_4(x)$ as our index function with $x_a = 1.25$. From Tables VI and II, therefore, we have

$$y_0 = \frac{x_a^2\sqrt{1 - x^2}}{H_y(x_a^2 - x^2)}; \quad x_a = 1.25, \tag{889}$$

or

$$y_0 = \sqrt{\frac{z_b}{z_a}} = \frac{j\omega\sqrt{\omega_1^2 - \omega^2}}{H_y(\omega_a^2 - \omega^2)}; \quad \omega_a = \frac{\omega_1}{x_a} = 0.8\omega_1. \tag{889a}$$

For the characteristic impedance function we shall suppose that $F_4(x)$ is also sufficient. With $x_2 = 1.2$, Fig. 158 shows that the tolerance is slightly more than 4 per cent with a coverage of nearly 90 per cent. Tables VI and I then give us the normalized characteristic impedance function either as

$$z_0 = \frac{x_2^2\sqrt{1 - x^2}}{H_z(x_2^2 - x^2)}; \quad x_2 = 1.20, \tag{890}$$

or

$$z_0 = \sqrt{z_b z_a} = \frac{\omega_2^2\sqrt{\omega_1^2 - \omega^2}}{H_z\omega_1(\omega_2^2 - \omega^2)}; \quad \omega_2 = \omega_1 x_2 = 1.20\omega_1. \tag{890a}$$

Note that the factor H has different numerical values in the functions y_0 and z_0 because the tolerances are different. By (881) we have for the problem discussed here

$$\left.\begin{aligned} H_y &= 1.02 \\ H_z &= 1.042 \end{aligned}\right\}.$$

With these we then find for the reactances z_a and z_b

$$\left.\begin{aligned} z_a &= \frac{1.410\omega_1(\omega_a^2 - \omega^2)}{j\omega(\omega_2^2 - \omega^2)} \\ z_b &= \frac{1.355\omega_1 j\omega(\omega_1^2 - \omega^2)}{(\omega_a^2 - \omega^2)(\omega_2^2 - \omega^2)} \end{aligned}\right\}. \tag{891}$$

The points at which minimum attenuation occur are given by (879) in terms of the variable x of the index function. These are in the present example

$$x = k\,sn\frac{\nu K}{2}; \quad \nu = 0, 1, 2,$$

and by means of tables we find for $k = 0.825$

$$x = 0;\ x = 0.659;\ x = 0.825.$$

The corresponding frequencies are then given by (888a), thus

$$\omega = 1.212\omega_1;\ \omega = 1.517\omega_1;\ \omega = \infty, \tag{892}$$

in which the order is reversed. The first of these is that frequency at which the minimum is first reached, as pointed out above.

The points of infinite attenuation are calculated from (882). They are[1]

$$x = k\, sn \frac{K}{4} = 0.38975;\ x = k\, sn \frac{3K}{4} = 0.79181,$$

and the corresponding frequencies become

$$\omega = 1.2629\omega_1;\ \omega = 2.5657\omega_1. \tag{892a}$$

It is interesting to note that the design so far is perfectly general with regard to the cut-off frequency ω_1 and the resistance R which the filter is to work into and out of. The component reactances Z_a and Z_b of the lattice are simply the expressions (891) multiplied by R. Their final realization then may be carried out easily by Foster's or Cauer's methods as discussed in Chapter V. There is no need of doing this here, the less so since the lattice itself is not the desirable structure for physical realization. The process of finding a ladder development for the network defined by the reactances (891) is discussed in a subsequent section. This process may be carried through for an arbitrary ω_1 more easily if we again use the substitution $x = \omega/\omega_1$. Thus we have with the numerical parameter values in (889a) and (890a)

$$\left.\begin{aligned} z_a &= \frac{1.410(0.64 - x^2)}{jx(1.44 - x^2)} \\ z_b &= \frac{1.355jx(1 - x^2)}{(0.64 - x^2)(1.44 - x^2)} \end{aligned}\right\}, \tag{891a}$$

where by (892a) the points of infinite attenuation correspond to

$$x = 1.2629 \text{ and } x = 2.5657. \tag{892b}$$

These are the roots of the equation $z_a - z_b = 0$, as is also clear from

[1] In the process of finding an equivalent ladder development for the lattice (which is discussed below), these frequencies, which correspond to points where $z_a = z_b$, must be known with considerable precision. Hence it is important that they be carefully evaluated and checked by substituting into the equation $z_a = z_b$. Since the existing tables of elliptic functions are frequently not precise enough, the values must be adjusted subsequently by a cut-and-try process. This was done in the present numerical case.

the fact that infinite attenuation is due to a balanced condition in the lattice (bridge).

Fig. 164 illustrates the reactances (891a) and serves to clarify some of the points brought out above.

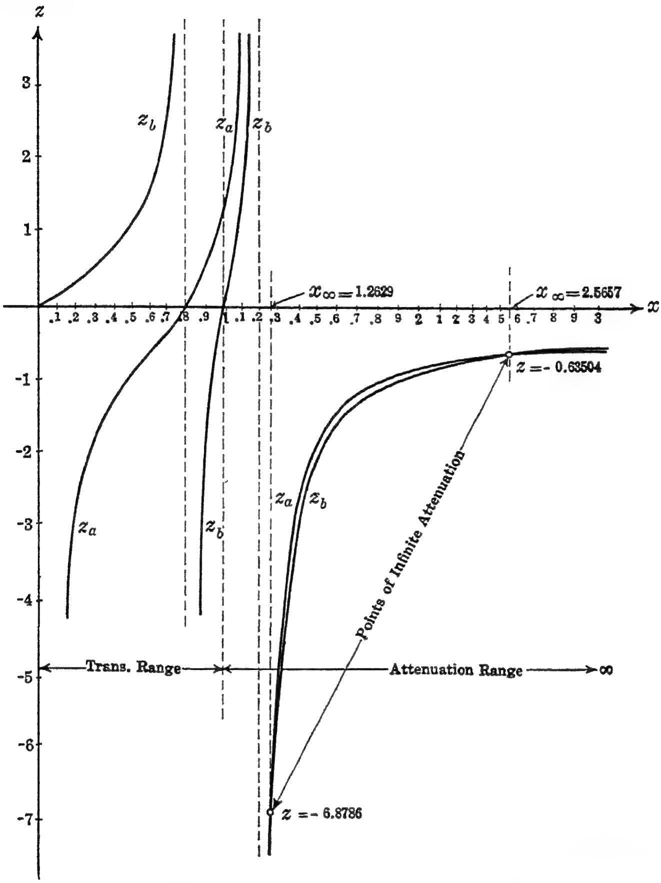


FIG. 164.—Behavior of the reactances (891a) versus the frequency variable $x = \omega/\omega_1$ for the low-pass lattice filter whose index and characteristic impedance functions are given by eqs. (889a) and (890a).

As a second example we choose the design of a band-pass filter whose theoretical cut-off frequencies are to lie at say 45,000 and 55,000 cycles per second, respectively, and for which the minimum attenuation shall again be 40 db, the latter to be reached at points not more than 500 cycles above and below the cut-offs. The transformations to the vari-

able x for the index function are given in Table V under the columns headed "Table IV." For the odd or even functions these are

$$x = \frac{w_1\omega_0}{\omega^2 - \omega_0^2} \text{ and } x = \frac{w_1\omega}{\omega^2 - \omega_0^2}$$

respectively, with due regard to the proper expressions for ω_0 and w_1 in the two cases. Substituting the frequencies $2\pi \cdot 44{,}500$ and $2\pi \cdot 55{,}500$ into these we find for odd functions the corresponding x-values

$$x = -0.918 \text{ and } x - 0.901,$$

while for even functions we have

$$x = -0.899 \text{ and } x = 0.917.$$

Since the larger of each pair of values determines the necessary coverage in either case, we see that for even or odd functions we shall need say $k = 0.92$ in order safely to meet the required sharpness of cut-off.

Hence the index function must be chosen to meet the requirements of 2 per cent tolerance with 8 per cent of the band not covered. Turning to Figs. 159 and 160 we see that F_5 fulfils our needs with a small margin in ϵ and $(1 - k)$ for the parameter values

$$\left.\begin{aligned} x_a &= 1.093 \\ x_b &= 2.230 \end{aligned}\right\}. \tag{893}$$

For the characteristic impedance function let us say that a tolerance of 4 per cent over 90 per cent is specified. Fig. 158 shows that this combination of values is just barely beyond the ability of $F_4(x)$. Since $F_5(x)$ is considerably better than required, however, we shall compromise on $F_4(x)$ with the parameter value

$$x_2 = 1.200. \tag{894}$$

The function y_0 is thus *number 5* in Table IV,[1] and z_0 is *number 4* of Table III, i.e.,

$$y_0 = \frac{(\omega_{-a}^2 - \omega^2)(\omega_b^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{\sqrt{\omega_{-1}^2 - \omega^2}(\omega_{-b}^2 - \omega^2)(\omega_a^2 - \omega^2)}, \tag{895}$$

and

$$z_0 = \frac{j\omega w_2^2\sqrt{(\omega_{-1}^2 - \omega^2)(\omega_1^2 - \omega^2)}}{Hw_1(\omega_{-2}^2 - \omega^2)(\omega_2^2 - \omega^2)}. \tag{896}$$

[1] Note that for this case the reduction to $F_5(x)$ by means of the transformation in Table V is exact.

For the calculation of the critical frequencies in these expressions we have, according to Table V, for the odd functions

$$\left.\begin{aligned} \omega_\nu &= \omega_0\sqrt{1+\frac{w_1}{\omega_0 x_\nu}} \\ \omega_{-\nu} &= \omega_0\sqrt{1-\frac{w_1}{\omega_0 x_\nu}} \end{aligned}\right\}, \tag{897}$$

in which

$$\omega_0 = 2\pi \cdot 50{,}250;\ w_1 = 2\pi \cdot 9{,}950, \tag{897a}$$

and for the even functions

$$\left.\begin{aligned} \omega_\nu &= \sqrt{\omega_0^2+\frac{x_\nu^2 w_1^2}{4}}+\frac{x_\nu w_1}{2} \\ \omega_{-\nu} &= \sqrt{\omega_0^2+\frac{x_\nu^2 w_1^2}{4}}-\frac{x_\nu w_1}{2} \end{aligned}\right\}, \tag{898}$$

in which

$$\omega_0 = 2\pi \cdot 49{,}750;\ w_1 = 2\pi \cdot 10{,}000. \tag{898a}$$

With (893) and (894) we thus have

$$\left.\begin{aligned} \omega_{-1} &= 2\pi \cdot 45{,}000\ ;\ \omega_1 = 2\pi \cdot 55{,}000 \\ \omega_{-2} &= 2\pi \cdot 44{,}100\ ;\ \omega_2 = 2\pi \cdot 56{,}100 \\ \omega_{-a} &= 2\pi \cdot 45{,}500\ ;\ \omega_a = 2\pi \cdot 54{,}600 \\ \omega_{-b} &= 2\pi \cdot 47{,}900\ ;\ \omega_b = 2\pi \cdot 52{,}400 \end{aligned}\right\}. \tag{899}$$

The exact tolerance in z_0, according to Fig. 158, is 4.2 per cent, so that $H = 1.042$. Since $w_2^2/w_1 = w_1x_2^2 = 2\pi\cdot 10^4\cdot 1.44$, the quotient and product of (896) and (895), respectively, yield

$$\left.\begin{aligned} z_a &= \frac{2\pi\cdot 10^4\cdot 1.382 j\omega(\omega_{-1}^2-\omega^2)(\omega_{-b}^2-\omega^2)(\omega_a^2-\omega^2)}{(\omega_{-2}^2-\omega^2)(\omega_{-a}^2-\omega^2)(\omega_b^2-\omega^2)(\omega_2^2-\omega^2)} \\ z_b &= \frac{2\pi\cdot 10^4\cdot 1.382 j\omega(\omega_{-a}^2-\omega^2)(\omega_b^2-\omega^2)(\omega_1^2-\omega^2)}{(\omega_{-2}^2-\omega^2)(\omega_{-b}^2-\omega^2)(\omega_a^2-\omega^2)(\omega_2^2-\omega^2)} \end{aligned}\right\}. \tag{900}$$

These multiplied by R are the component reactances Z_a and Z_b, the networks for which may be found by Foster's or Cauer's methods.

The points of minimum and infinite attenuation are calculated from (886) and (887), respectively. For the minima we have

$$x = k\,sn\frac{\nu K}{5};\ \nu = 1, 3, 5.$$

With $k = 0.924$, which is the exact value for the parameters (893), this gives

$$x = 0.4155;\ x = 0.8460;\ x = 0.9240. \tag{901}$$

Substituting these and (897a) into (897) we find the frequency pairs

$$\left.\begin{array}{c}36{,}350;\ 61{,}020\\ 43{,}980;\ 55{,}820\\ 44{,}550;\ 55{,}360\end{array}\right\} \tag{901a}$$

respectively, of which the last represents those frequencies nearest the cut-offs at which the minimum attenuation is first reached.

For the infinite maxima we have

$$x = k\,sn\frac{2\nu K}{5};\ \nu = 0, 1, 2,$$

which gives

$$x = 0;\ x = 0.6998;\ x = 0.9078. \tag{902}$$

Again using (897) and (897a), the corresponding frequencies are found to be

$$\left.\begin{array}{c}\text{infinity}\\ 42{,}540;\ 56{,}900\\ 44{,}420;\ 55{,}470\end{array}\right\}. \tag{902a}$$

Note here that the value $x = 0$ does not give a pair of frequencies as do the other values, but merely the point infinity. At the origin the attenuation is finite but greater than the minimum value.

9. Bode's method of parameter determination. The ideas involved in this process have much in common with those pertaining to the design of non-dissipative artificial lines or all-pass linear phase shift networks by means of the lattice structure, as discussed in section 2 of Chapter VII. This is due to the fact that Bode approaches the problem with the view toward making the phase characteristic of the filter in its transmission range as linear as possible. If we make this stipulation, we evidently have for such a range

$$\gamma = j\gamma_2 = j\omega t_d, \tag{903}$$

where t_d is the desired constant slope of the phase function. The corresponding index function then becomes[1]

$$y_0 = \tanh\frac{\gamma}{2} = j\tan\frac{\omega t_d}{2}, \tag{904}$$

which has the form of a reactance function with uniformly spaced zeros and poles.

[1] According to (846) and (853) the index function is the reciprocal of tanh $\gamma/2$. Since interchanging Z_a and Z_b has no effect other than that of adding π to the phase, this is of no particular importance.

As pointed out previously, the Cauer method of parameter determination does not lead to a linear phase characteristic but rather to one which is decidedly concave. This may be seen by considering, for example, a low-pass case. According to Table II, the index function has the form[1]

$$y_0 = \frac{j\omega(\omega_{n-1}^2 - \omega^2)(\omega_{n-3}^2 - \omega^2) \cdots (\omega_3^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{(\omega_n^2 - \omega^2)(\omega_{n-2}^2 - \omega^2) \cdots (\omega_4^2 - \omega^2)(\omega_2^2 - \omega^2)}, \tag{905}$$

for n-even. Since in this case $x_\nu = \omega_1/\omega_\nu$, equation (878) gives

$$\omega_\nu = \omega_1\, sn\left(\frac{n - \nu + 1}{n}\right)K; \; (\nu = 1, 2, 3, \cdots n), \tag{905a}$$

i.e., the critical frequencies are spaced according to the amplitudes of the sn-function for equal increments in its argument. Since the phase increases by $\pi/2$ for each increment in ω from one critical frequency to the next, it follows that at the critical frequencies we have

$$\gamma_2 = \frac{n\pi}{2}\, sn^{-1}\frac{\omega}{\omega_1} K; \; (\omega = \omega_\nu).$$

Thus the general character of γ_2 is seen to follow an inverse sn-function, which is quite concave, more so than the inverse trigonometric sine.

To avoid this, Bode recognizes that a uniform spacing of the critical frequencies is essential over that region throughout which a linear phase characteristic is desired. Although this is a necessary condition, it is quite evidently not sufficient because of the finite extent of the region. Using the infinite product expansion of the tangent function we have for the transmission range

$$y_0 = \frac{j\omega t_d}{2} \cdot \frac{\left(1 - \frac{\omega^2}{\omega_2^2}\right)\left(1 - \frac{\omega^2}{\omega_4^2}\right)\cdots}{\left(1 - \frac{\omega^2}{\omega_1^2}\right)\left(1 - \frac{\omega^2}{\omega_3^2}\right)\cdots}, \tag{904a}$$

where the critical frequencies are defined by[2]

$$\omega_\nu t_d = \nu\pi. \tag{904b}$$

[1] The factor H is not material to the present discussion. The numerical subscripts 2, 3, 4 . . . are here more suitable than the literal ones a, b, c, . . . in Table II.

[2] Note that the critical frequencies in expression (904a) do not follow the same order according to subscripts as those in expressions for y_0 in the preceding discussion of Cauer's method of parameter determination. Here ω_1 is the critical frequency nearest the origin, ω_2 the next, etc. This change in notation is necessary at this time since the discussion of Bode's method in terms of the former system of notation leads to extremely awkward expressions.

The exact realization of a linear phase characteristic with uniformly spaced critical frequencies thus requires a transmission range of infinite extent. Actually y_0 behaves like a reactance function only over a finite range and becomes real and positive beyond the cut-off point.

The actual situation may be pictured as in Fig. 165. Here the regions A and C represent the transmission and attenuation ranges, respectively, of a low-pass filter. B is a transition region throughout which the reactances z_a and z_b pass from the tendency $z_a z_b \triangleq 1$ to the tendency $z_a \triangleq z_b$. The theoretical cut-off frequency f_c lies somewhere within this transition range.

Suppose that the region A embraces the first n critical frequencies

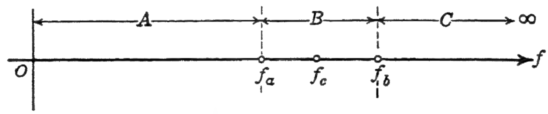


FIG. 165.—Division of the frequency spectrum for a low-pass filter into the practical transmission and attenuation ranges A and C, and a transition range B which includes the theoretical cut-off f_c.

of the expression (904a) and that we denote this portion of the infinite product by P and the remainder by Q so that we can write

$$y_0 = P \cdot Q, \tag{906}$$

where

$$P = \frac{j\omega t_d}{2} \cdot \frac{\left(1 - \frac{\omega^2}{\omega_2^2}\right)\left(1 - \frac{\omega^2}{\omega_4^2}\right) \cdots \left(1 - \frac{\omega^2}{\omega_{n-1}^2}\right)}{\left(1 - \frac{\omega^2}{\omega_1^2}\right)\left(1 - \frac{\omega^2}{\omega_3^2}\right) \cdots \left(1 - \frac{\omega^2}{\omega_n^2}\right)}, \tag{906a}$$

and

$$Q = \frac{\left(1 - \frac{\omega^2}{\omega_{n+1}^2}\right)\left(1 - \frac{\omega^2}{\omega_{n+3}^2}\right) \cdots}{\left(1 - \frac{\omega^2}{\omega_{n+2}^2}\right)\left(1 - \frac{\omega^2}{\omega_{n+4}^2}\right) \cdots}, \tag{906b}$$

which is arbitrarily written for *n-odd*. Our problem is now this: Whereas Q extends from the boundary frequency f_a (Fig. 165) to infinity, we must find such a function Q' whose critical frequencies occupy only a part of the region B ($f_a < f < f_c$), but which approximates Q within the region A sufficiently well to satisfy the requirement of linearity in γ_2. At the same time Q' must also approximate $1/P$ within the region C in order that y_0 may there approximate *unity* and thus fulfil the attenuation requirement of the filter. The success of Bode's method is due to the fact that such a function Q' can be found.

In order to satisfy these requirements even approximately, Q' evidently must be real throughout A and imaginary throughout C. This suggests that it contain an irrational factor of the form

$$\sqrt{1 - \frac{\omega^2}{\omega_c^2}}, \tag{907}$$

where the critical frequency ω_c may lie anywhere within the region B and thus become the theoretical cut-off. In fact, the familiar form of y_0, according to Table II, makes such a factor an essential part of the function Q'. Again recalling our discussion of artificial line design in Chapter VII, and referring to Fig. 97, page 265, we see that the irrational factor alone already approximates the desired Q within the region A fairly well when ω_c is properly chosen. This is also quite readily seen from (906*b*) which is evidently unity at $\omega = 0$ and varies uniformly from this value to zero at $\omega = \omega_{n+1}$.

For large values of ω, P, according to (906*a*), becomes

$$P \rightarrow \frac{t_d(\omega_1\omega_3 \cdots \omega_n)^2}{2j\omega(\omega_2\omega_4 \cdots \omega_{n-1})^2}, \tag{908}$$

while the irrational factor (907) approaches the function $j\omega/\omega_c$, which may very well approximate $1/P$. In fact, using (904*b*) and writing $\omega_c t_d = c\pi$, we find by equating the reciprocal of (908) to $j\omega/\omega_c$

$$c = \frac{\pi}{2}\left(\frac{1 \cdot 3 \cdot 5 \cdots n}{2 \cdot 4 \cdot 6 \cdots n-1}\right)^2, \tag{909}$$

which, for large n, approaches

$$c = n + \tfrac{1}{2} \tag{909a}$$

as may be seen from a few numerical trials. To a first approximation, therefore, our problem is solved by

$$Q' = \sqrt{1 - \frac{\omega^2}{\omega_{n+\frac{1}{2}}^2}}. \tag{910}$$

In order to obtain a better approximation in both the ranges A and C, a more elaborate form may be assumed for Q'. With the necessary form of y_0 in mind, this must be chosen as

$$Q' = \frac{\left(1 - \frac{\omega^2}{\omega_{p1}^2}\right)\left(1 - \frac{\omega^2}{\omega_{p3}^2}\right) \cdots \left(1 - \frac{\omega^2}{\omega_{pm-1}^2}\right)}{\left(1 - \frac{\omega^2}{\omega_{p2}^2}\right)\left(1 - \frac{\omega^2}{\omega_{p4}^2}\right) \cdots \sqrt{1 - \frac{\omega^2}{\omega_{pm}^2}}}, \tag{910a}$$

in which the irrational factor appears in the numerator or denominator according to whether m is *odd* or *even*. Incidentally, if n is even, Q' is given by the reciprocal of (910*a*).

By means of a rather lengthy manipulation process, the details of which we omit entirely, Bode[1] evaluates the critical frequencies ω_{p_s} for various values of m. The results are given in Table VIII for which the following substitution is used

$$p_s = n + 1 + \delta_s. \tag{911}$$

The critical frequencies are given by (904b) for $\nu = p_s$. The theoretical cut-off frequency of the filter is ω_{pm}.

TABLE VIII

m	δ_1	δ_2	δ_3	δ_4	δ_5
1	−0 5000				
2	−0 14645	+0 20711			
3	−0 05032	+0 67731	+0 95526		
4	−0 01897	+0 86157	+1 49180	+1 72252	
5	−0 00759	+0 93720	+1 75091	+2 30794	+2 50684

With each additional factor, Q' approximates Q in the transmission range, and $1/P$ in the attenuation range, with greater accuracy. If we write for the transmission range

$$Q' = Q(1 + \epsilon), \tag{912}$$

then ϵ becomes the decimal error by which the index function y_0 fails to equal its ideal expressed by (904). For small errors, this may easily be correlated with the resulting deviation from linearity in γ_2 in the following way. We have

$$y_0 = j \tan \frac{\gamma_2}{2}.$$

Taking the differential of both sides gives

$$dy_0 = \frac{j d\gamma_2}{2 \cos^2 \frac{\gamma_2}{2}},$$

[1] This is described in the patent specification already referred to. Briefly the process consists of representing the infinite product Q by means of gamma functions and these in turn by their Stirling series approximations valid for large n. After further changes of form, the expansion of Q is compared with that of Q' and coefficients of like powers equated. Out of these equations the parameters are finally evaluated. It should be noted that the numerical results of Table VIII thus obtained are valid for large n. For example, the resulting y_0-function for $m = 1$, which involves the half-spaced cut-off factor, approaches unity for $\omega \to \infty$ only for $n \to \infty$. For a finite n, $y_0 = 1$ at a finite frequency; i.e., the filter does have a point of infinite attenuation, although for a reasonably large n this occurs far from the cut-off.

and for a small ϵ it follows that

$$\frac{dy_0}{y_0} = \epsilon = \frac{d\gamma_2}{2 \sin \frac{\gamma_2}{2} \cos \frac{\gamma_2}{2}} = \frac{d\gamma_2}{\sin \gamma_2}.$$

Hence the corresponding deviation in γ_2 becomes, to a first approximation

$$d\gamma_2 = \epsilon \sin \omega t_d. \tag{913}$$

At the critical frequencies the deviation in γ_2 is zero as is to be expected. Between these, however, γ_2 deviates from a straight line in the form of

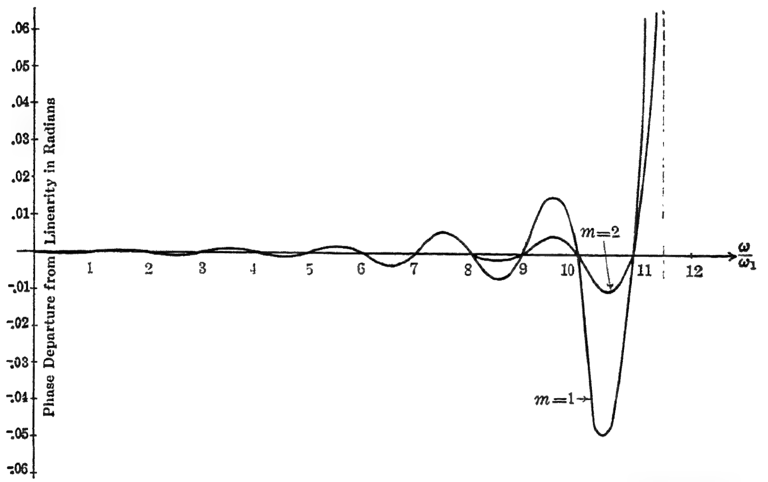


FIG. 166.—Departures from linearity in the phase function of a low-pass filter designed on the Bode basis with 11 factors in P and 1 or 2 factors in Q'.

a half-period of the sine-function with an amplitude equal to the error in y_0. With a constant ϵ, γ_2 would oscillate about a straight line with uniform amplitude. Actually the above approximation method results in an error ϵ which is zero for $\omega = 0$, increases very slowly over the region A (Fig. 165), and more rapidly from the end of this region up to the theoretical cut-off. Hence γ_2 is practically linear throughout A, but begins to show a slightly oscillatory character with gradually increasing amplitude toward the end of this region. Fig. 166 shows the departures of γ_2 from linearity for the cases $m = 1$ and $m = 2$, both with $n = 11$; Fig. 167 illustrates the corresponding attenuation curves, i.e., plots of γ_1.[1]

[1] Both figures taken from notes by H. W. Bode.

The significant feature of Bode's method of parameter determination is that it combines linearity in the phase characteristic over the transmission range with a rapidly rising attenuation characteristic in the attenuation range. The latter characteristic, in contrast to that resulting from Cauer's method of parameter determination, does not show infinite maxima with equal intervening minima, but rises toward infinity in a uniform manner, the rapidity of this rise depending upon the complexity of the function Q', i.e., upon the number of non-coincident[2] critical frequencies chosen within the transition range B. It should be recognized in this connection that although the improvement of both the phase and attenuation characteristics of the filter depend largely upon the number of factors chosen for Q', they also are

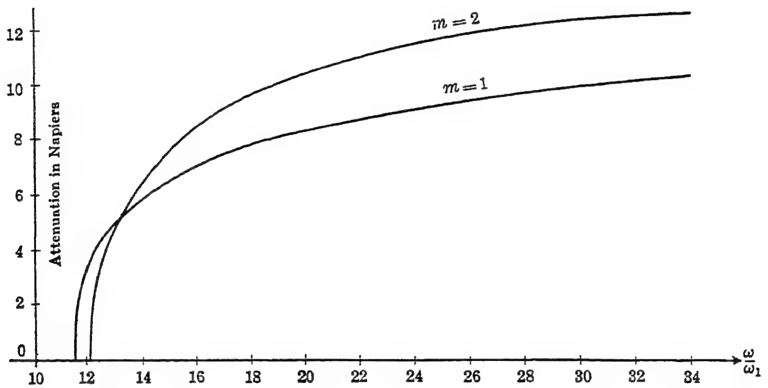


FIG. 167.—Attenuation functions corresponding to the low-pass filter designs whose phase departures are shown in Fig. 166.

affected by the number (n) of factors chosen for P. Thus, for a given total number of factors in $y_0 = P \cdot Q'$, the quality of the filter characteristics depends upon the allotment of these factors between P and Q'. The Figs. 166 and 167 show, however, that it is unnecessary for most practical designs to allot more than two factors to Q'. Furthermore it should be noted that the parameter values in Table VIII lead to a compromise between the quality of the resulting attenuation and phase functions and may need to be modified by trial to shift the resulting quality in favor of either the phase or the attenuation characteristic respectively as called for by the requirements of a specific design. This point is emphasized to some extent in the last section of this chapter.

[2] The first n critical frequencies of y_0 coincide with those of the tangent function. The remaining m critical frequencies are spoken of, therefore, as *non-coincident*.

The function y_0, within its approximation range, does not oscillate about unity, but approaches this value uniformly as illustrated by the curve B of Fig. 149. That is, whereas the Cauer method gives rise to real, simple roots of the equation $y_0 = 1$ within the attenuation range, these roots are *complex* in Bode's method. We shall see that, as a consequence of this, the resulting lattice cannot be completely developed in the ladder form, which is a disadvantage from the practical standpoint.

It is quite evident that the same method of parameter determination may also be applied to the characteristic impedance z_0, since it is the same kind of function as y_0 when a suitable change of variable is made use of. In order to be able to apply the content of Table V as such substitutions, it must be noted, of course, that ω_1 in that table refers to the cut-off frequency of the low- or high-pass classes. This must be identified with ω_{pm} in the above analysis, while the critical frequencies $\omega_1, \omega_2, \ldots \omega_n, \omega_{p1}, \ldots \omega_{pm-1}$ must be identified with $\omega_l, \ldots \omega_c, \omega_b, \omega_a$, respectively, where ω_l is written for the lowest critical frequency in the functions of Table II and hence replaces ω_1 in the Bode analysis. This seeming confusion in the notation is difficult to avoid in a side-by-side discussion of Bode's and Cauer's methods, but should be easily overcome by the reader who has studied the whole situation sufficiently to reach clarity on the chief points at issue in both methods.

For example, suppose we consider a y_0-function on the Bode basis with three uniformly spaced critical frequencies and a Q'-function of the form (910), i.e., the general form (910a) for $m = 1$. Thus we would have

$$y_0 = \frac{j\omega t_d}{2} \frac{\left(1 - \dfrac{\omega^2}{\omega_2^2}\right)\sqrt{1 - \dfrac{\omega^2}{\omega_{p1}^2}}}{\left(1 - \dfrac{\omega^2}{\omega_1^2}\right)\left(1 - \dfrac{\omega^2}{\omega_3^2}\right)}, \tag{914}$$

or in the notation of Table V

$$\begin{aligned} y_0 &= \frac{j\omega t_d}{2} \frac{\left(1 - \dfrac{\omega^2}{\omega_b^2}\right)\sqrt{1 - \dfrac{\omega^2}{\omega_1^2}}}{\left(1 - \dfrac{\omega^2}{\omega_c^2}\right)\left(1 - \dfrac{\omega^2}{\omega_a^2}\right)} \\ &= \frac{j\omega\sqrt{\omega_1^2 - \omega^2}(\omega_b^2 - \omega^2)\omega_a^2\omega_c^2 t_d}{(\omega_a^2 - \omega^2)(\omega_c^2 - \omega^2)2\omega_1\omega_b^2}, \end{aligned} \tag{914a}$$

which is identified with function 8 of Table II if we put

$$H = \frac{2\omega_1\omega_b^2}{t_d\omega_a^2\omega_c^2}. \tag{915}$$

Applying (904b) we have

$$\left.\begin{aligned} \omega_1 &\to \omega_c = \frac{\pi}{t_d} \\ \omega_2 &\to \omega_b = \frac{2\pi}{t_d} \\ \omega_3 &\to \omega_a = \frac{3\pi}{t_d} \\ \omega_{p1} &\to \omega_1 = \frac{7\pi}{2t_d} \end{aligned}\right\}. \tag{915a}$$

Using these in (915) we find

$$H = \frac{4}{\pi} \cdot \frac{7}{9} = 0.991.$$

This means, incidentally, that the $y_0 = 1$ equation will have a real root at some finite frequency and hence the attenuation function will show an infinite maximum. This is due to the fact that the result (909a) follows from (909) only as n becomes very large. In that case H equals unity. In the present example $n = 3$, so that the half-spaced critical frequency in Q' is not quite correct. This does no particular harm, however.

The correlation with our former notation having been established, the frequency transformations of Table V may be used again, by means of which the Bode method is easily extended from the low-pass class (to which the preceding discussion is entirely restricted) to that of the band pass.[1] This point brings us to an interesting and useful process which may be used with equal effectiveness in the design of filters on the Cauer basis. Namely, it is possible to design a filter of any of the four common classes discussed here, as a low-pass filter first, and then transform to the desired class afterward.

[1] Here it should be noted that the phase characteristic of the resulting band-pass filter will have the same degree of linearity as that of the low-pass only when the band width is small compared with the mean frequency. In mathematical terms the following approximation in the substitution for the band-pass filter

$$x = \frac{\omega^2 - \omega_0^2}{w_1\omega} \cong \frac{2(\omega - \omega_0)}{w_1}$$

is valid for the range $-1<x<1$ only so long as $w_1/2\omega_0$ is a negligible quantity. When this is not the case, the parameters in y_0 must subsequently be readjusted by a cut-and-try process. It should be noted also in this connection that the phase characteristic in the transmission range of a high-pass filter or in the upper transmission band of a band-elimination filter *cannot* be linear on account of the infinite extent of these bands and the finite total phase shift. In these cases an approximately linear characteristic can be obtained over a finite portion of the transmission bands, although the parameter determination for these is not given simply by means of frequency transformations applied to a low-pass design.

For example, suppose we have designed a low-pass filter with cut-off at ω_1 and have found the reactance functions z_a and z_b which are written similar to (891a) in terms of the variable $x = \omega/\omega_1$. A band-pass filter with the same relative attenuation, phase, and characteristic impedance functions is immediately obtained by means of the transformation

$$x = \frac{\omega^2 - \omega_0^2}{w_1\omega} \tag{916}$$

of Table V, where $\omega_0 = \sqrt{\omega_1\omega_{-1}}$, $w_1 = \omega_1 - \omega_{-1}$, and ω_1, ω_{-1} are the theoretical cut-off frequencies.

To carry out this transformation we note that whereas an impedance

$$njx = \frac{nj\omega}{\omega_1} \tag{917}$$

represents a coil with an inductance of n/ω_1 henries, and an impedance

$$\frac{1}{njx} = \frac{\omega_1}{nj\omega} \tag{917a}$$

represents a capacitance of n/ω_1 farads, an impedance

$$njx = \frac{nj(\omega^2 - \omega_0^2)}{w_1\omega} = \frac{1 - L_1C_1\omega^2}{jC_1\omega} \tag{918}$$

represents a coil and condenser in series with

$$L_1 = \frac{n}{w_1};\ \omega_0 = \frac{1}{\sqrt{L_1C_1}};\ C_1 = \frac{w_1}{n\omega_0^2}, \tag{918a}$$

while an impedance

$$\frac{1}{njx} = \frac{w_1\omega}{nj(\omega^2 - \omega_0^2)} = \frac{jL_2\omega}{1 - L_2C_2\omega^2} \tag{919}$$

represents a coil and condenser in parallel with

$$C_2 = \frac{n}{w_1};\ \omega_0 = \frac{1}{\sqrt{L_2C_2}};\ L_2 = \frac{w_1}{n\omega_0^2}. \tag{919a}$$

Hence, if, in the network of the low-pass filter, we replace each coil (having a value of n/ω_1 henries) by a resonant component with constants determined according to (918a), and each condenser (having a value of n/ω_1 farads) by an anti-resonant component with constants determined according to (919a), the resulting structure will be the desired band-pass filter. The reader should note that this method is not restricted to filters of the type discussed in the present chapter, but holds generally and may be applied to the conventional filter as well.

Similarly the band-elimination filter is obtained from the low-pass by replacing coils by anti-resonant components and condensers by resonant components, or by reciprocating the components of the band-pass; the high-pass results from the low-pass by reciprocating the elements of the low-pass.

Another interesting point follows from these considerations which may be useful in connection with the band-pass filter obtained by this method. Equations (918a) and (919a) show that the band width depends only upon L_1 and C_2. Hence if it is desired to keep this constant but shift the location of the band, it is merely necessary to vary all the condensers C_1 and all the coils L_2 by proportionate amounts; i.e., these may be controlled from a single shaft.

It should be noted in this connection also that this process of designing band-pass or band-elimination filters from the low-pass, limits us to the choice of even functions for y_0 and z_0. The Bode method may, incidentally, also be carried out with odd functions for the **B.P.** and **B.E.** classes, although it is inconvenient to formulate general relationships for this purpose.

Finally it might be mentioned that, in the design of a filter, it is quite possible to determine the y_0-function on the Bode basis and the z_0-function on Cauer's, since these functions are entirely independent. In view of the fact that Cauer's process is somewhat more economical on the basis of a given tolerance and coverage for z_0, such a procedure may seem to be in order under certain circumstances.

10. Ladder developments and other lattice equivalents. It has been pointed out earlier that the lattice structure itself is an impractical form for realizing the properties predicted by the design, primarily because the high attenuation values are due to a Wheatstone bridge balance, the attainment of which requires extreme precision in the adjustment of parameter values as well as careful attention to other matters incident to bridge measurement technique such as balance with respect to the power source as well as to ground, etc. These matters impose a fairly definite upper limit upon the attenuation which may be realized from a single lattice structure.

Since the ladder form of network is relatively free from practical difficulties of this sort, it offers a more suitable means for the realization of the predicted behavior provided the given lattice can be so developed. Bode has shown that, if the parameters in both y_0 and z_0 are determined on the Cauer basis, such a development is always possible. The manner in which this may be accomplished will now be taken up.

It will be recalled[1] that the reactances of the symmetrical lattice are given by

$$\left.\begin{aligned} Z_a &= Z_0 \tanh \frac{\gamma}{2} \\ Z_b &= Z_0 \coth \frac{\gamma}{2} \end{aligned}\right\}, \tag{920}$$

and also that for the general dissymmetrical two-terminal pair[2]

$$\left.\begin{aligned} z_{11} &= Z_{I1} \coth \theta \\ \frac{1}{y_{11}} &= Z_{I1} \tanh \theta \end{aligned}\right\}, \tag{920a}$$

where Z_{I1} is its image impedance for end 1 and θ the corresponding image propagation function. Since two dissymmetrical networks placed back-to-back correspond to one symmetrical network whose propagation function equals twice the image propagation function of the individual networks, and whose characteristic impedance equals the image impedance for the non-adjacent ends, it follows that Z_a may be looked upon as the $1/y_{11}$ (short-circuit impedance) and Z_b as the z_{11} (open-circuit impedance) of *half* the original network. In other words, if we imagine the original lattice, or any equivalent thereof, bisected[3] then Z_a and Z_b are the short-circuit and open-circuit reactances, respectively, of either dissymmetrical half viewed from its original terminals.

The method of ladder development depends upon the fact that, if the open- and short-circuit impedances possess a common additive term, then this may be taken out as a series branch on the input side of the network, while if the open- and short-circuit admittances contain a common additive term, this may be taken out as a shunt branch. By alternating between these processes a ladder form is evidently obtained. To illustrate, suppose we have

$$\left.\begin{aligned} Z_a &= aZ_1 + Z_{a2} \\ Z_b &= bZ_1 + Z_{b2} \end{aligned}\right\}. \tag{921}$$

Here either aZ_1 or bZ_1 may be removed, depending upon whether $a > b$ or $b > a$. If $b > a$, we then have for the remaining network

$$\left.\begin{aligned} Z_a' &= Z_{a2} = \frac{1}{Y_a'} \\ Z_b' &= (b - a)Z_1 + Z_{b2} = \frac{1}{Y_b'} \end{aligned}\right\}. \tag{921a}$$

[1] See eq. (432), p. 180.

[2] See eqs. (413) and (414), p. 175.

[3] The bisection of the lattice itself must be interpreted in the sense shown in Brune's "Note on Bartlett's Bisection Theorem" (ref. on p. 439); see particularly his Fig. 6.

This is illustrated in Fig. 168.

Now suppose that

$$\left.\begin{aligned} Y_a' &= a'Y_1 + Y_{a2}' \\ Y_b' &= b'Y_1 + Y_{b2}' \end{aligned}\right\}, \tag{922}$$

where $b' > a'$. Then the admittance $a'Y_1$ may be removed as a shunt branch, leaving

$$\left.\begin{aligned} Y_a'' &= Y_{a2}' = \frac{1}{Z_a''} \\ Y_b'' &= (b' - a')Y_1 + Y_{b2}' = \frac{1}{Z_b''} \end{aligned}\right\}. \tag{922a}$$

The result up to this point is illustrated in Fig. 169. From here on, the same cycle of events is repeated until the remainder is either a single series or shunt branch.

The success of the development depends upon the ability at each step to remove a common term in the pair of impedance or admittance

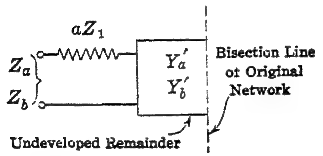


FIG. 168.—Schematic representation of the first step in the ladder development according to the relations (921) and (921a).

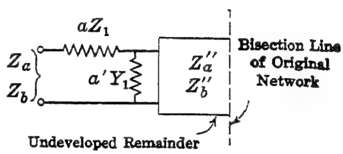


FIG. 169.—The ladder development of Fig. 168 carried one step farther by removing a shunt admittance according to the relations (922) and (922a).

functions, in our case reactances or susceptances. Recalling Foster's reactance theorem, this will evidently be so if the functions have one or more common poles, when their partial fraction expansions will contain the necessary common terms. The removal (in whole or part) of these common poles must leave functions with one or more common zeros, which, upon inversion, again become common poles for the functions of the remainder.

The details of this method can be described best by means of an illustrative numerical example. For this purpose let us return to our low-pass filter design in section 8, the normalized reactances for which are given by (891a) where $x = \omega/\omega_1$. Expanding these in partial fractions we get

$$\left.\begin{aligned} z_a &= \frac{0.62666}{jx} + \frac{0.78333jx}{1.44 - x^2} \\ z_b &= \frac{0.60975jx}{0.64 - x^2} + \frac{0.74525jx}{1.44 - x^2} \end{aligned}\right\}. \tag{923}$$

These evidently have a common pole at $x = 1.20$. Hence we can remove the series reactance branch

$$z_1 = \frac{0.74525jx}{1.44 - x^2} = \frac{1}{1.3418jx + \frac{1}{0.51754jx}}, \tag{924}$$

leaving

$$\left.\begin{aligned} z_a' &= \frac{0.62666}{jx} + \frac{0.03808jx}{1.44 - x^2} = \frac{0.66474(1.3575 - x^2)}{jx(1.44 - x^2)} \\ z_b' &= \frac{0.60975jx}{0.64 - x^2} \end{aligned}\right\}, \tag{923a}$$

or, inverting and again expanding in partial fractions,

$$\left.\begin{aligned} y_a' &= \frac{0.12411jx}{1.3575 - x^2} + 1.5043jx \\ y_b' &= \frac{1}{0.95312jx} + 1.6393jx \end{aligned}\right\}. \tag{923b}$$

Since the remaining reactances (923a) have a common zero at infinity, the susceptances (923b) have a common pole there. These susceptances are sketched in Fig. 170 in which the curves and scales are purposely distorted in order that the important characteristics may stand out more prominently. The points at which the susceptances are equal correspond to the infinite maxima in the attenuation function of the filter and, therefore, are known from (892b).

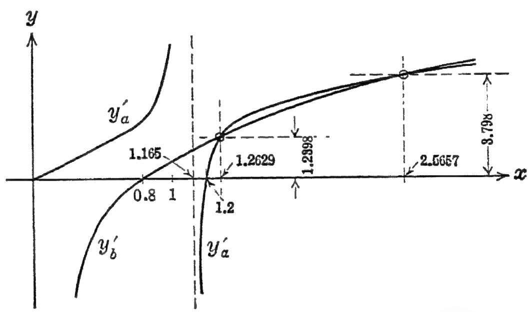


FIG. 170.—Curves showing the significant features in the behavior of the susceptance functions (923b).

The next step in the procedure must be carried out in a somewhat different fashion. Namely, if we remove all of the pole at infinity in y_a', the remaining susceptances have no common zero, and hence the process could be carried no farther. However, if we remove only that part of the pole corresponding to either one of the common values, i.e., take out a susceptance ajx which also passes through one of the points $y_a' = y_b'$, then the remaining susceptances will evidently have a common zero at that point so that we will be assured of being able to proceed in the subsequent step of the development.

Since either one or the other of the common values must be chosen arbitrarily, we will first try taking out the value j 1.2398 which occurs at $x = 1.2629$, corresponding to a susceptance

$$y_1 = \frac{1.2398jx}{1.2629} = 0.98171jx, \tag{924a}$$

leaving

$$\left.\begin{aligned} y_a'' &= \frac{0.12411jx}{1.3575 - x^2} + 0.5226jx = \frac{0.5226jx(1.5952 - x^2)}{1.3575 - x^2} \\ y_b'' &= \frac{1}{0.95312jx} + 0.6576jx = \frac{1.5952 - x^2}{1.5207jx} \end{aligned}\right\}. \tag{923c}$$

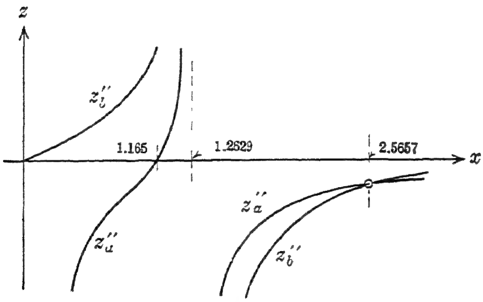


FIG. 171.—Curves showing the significant features in the behavior of the reactance functions (923d).

Inverting this remainder and again expanding in partial fractions gives

$$\left.\begin{aligned} z_a'' &= \frac{1}{0.61411jx} + \frac{0.28513jx}{1.5952 - x^2} \\ z_b'' &= \frac{1.5207jx}{1.5952 - x^2} \end{aligned}\right\}. \tag{923d}$$

These are sketched in Fig. 171. Here the only thing to be done is to remove the pole at $x = 1.2629$ entirely from z_a'', i.e., take out

$$z_2 = \frac{0.28513jx}{1.5952 - x^2} = \cfrac{1}{3.5072jx + \cfrac{1}{0.17874jx}}, \tag{924b}$$

leaving

$$\left.\begin{aligned} z_a''' &= \frac{1}{0.61411jx} \\ z_b''' &= \frac{1.2356jx}{1.5952 - x^2} \end{aligned}\right\}, \tag{923e}$$

or

$$\left.\begin{aligned} y_a''' &= 0.61411jx \\ y_b''' &= \frac{1}{0.77457jx} + 0.80932jx \end{aligned}\right\}, \tag{923f}$$

which again have a pole at infinity because z_a'' and z_b'' had a zero there. Taking out the susceptance

$$y_2 = 0.61411jx, \tag{924c}$$

leaves

$$y_a^{\text{iv}} = 0; \; y_b^{\text{iv}} = \frac{1}{0.77457jx} + 0.19521jx. \tag{923g}$$

This remaining undeveloped portion is more easily set down by regarding y_a^{iv} and y_b^{iv} as the component susceptances of a symmetrical lattice forming the central part of the entire development. Since $y_a^{\text{iv}} = 0$, this lattice consists only of its cross-arms which, therefore, may be combined into one susceptance of half the value of y_b^{iv}. The crossing of the arms merely amounts to a shift of π radians in the phase and hence may be omitted.[1] With the reactance and susceptance values (924), (924a), (924b), (924c), and the remainder (923g), the final ladder development is given by the network of Fig. 172. The dotted line indicates the center of the structure which, of course, is symmetrical.

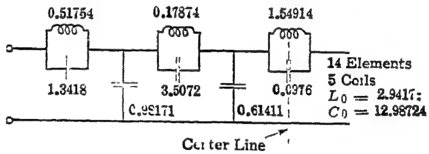


FIG. 172.—Ladder development of the low-pass lattice filter whose component reactances are shown in Fig. 164. The parameter values in this development are obtained from the relations (924), (924a), (924b), (924c), and (923g).

The total number of elements necessary for this realization is 14; the number of coils 5. From the forms of z_a and z_b as given by (923), the lattice form evidently involves 14 elements and 6 coils. From the standpoint of the number of elements alone, the ladder development obviously exhibits no advantage. This is no fair basis upon which to discriminate, however, since practical considerations as well as tolerances on the element values are important items in affecting cost. On this basis the ladder development is much to be preferred.

Incidentally, the actual inductance and capacitance values in a specific case depend upon ω_1 (the cut-off) and R (the resistance which

[1] It is interesting to note that the distinction between the subscripts a and b does not appear until this last step when it merely amounts to a switching of the terminals. This checks with what was said in section 4.

the structure is to work into and out of). Recalling that $x = \omega/\omega_1$ and that the actual lattice reactances are z_a and z_b multiplied by R, it is clear that all inductance values in Fig. 172 are to be multiplied by R/ω_1 and all capacitance values by $1/R\omega_1$. Also it again may be mentioned that a band-pass filter with the same relative propagation and characteristic impedance functions is immediately obtained by replacing coils by resonant components and condensers by anti-resonant components according to the relations (916) to (919*a*).

An alternative development may be attempted by returning to the point in the previous one at which the functions (923*b*), illustrated in Fig. 170, were arrived at. Instead of taking out the smaller of the common values we shall take out the larger ($y_a' = y_b' = j\,3.7980$) which occurs at $x = 2.5657$, corresponding to a susceptance

$$y_1 = \frac{3.7980jx}{2.5657} = 1.4803jx, \tag{925}$$

leaving

$$\left.\begin{aligned} y_a'' &= \frac{0.12411jx}{1.3575 - x^2} + 0.0240jx = \frac{0.0240jx(6.583 - x^2)}{1.3575 - x^2} \\ y_b'' &= \frac{1}{0.95312jx} + 0.1590jx = \frac{6.583 - x^2}{6.2893jx} \end{aligned}\right\}. \tag{926}$$

Thus we have created a common zero at $x = 2.5657$. The partial fraction expansions of the corresponding reactances read

$$\left.\begin{aligned} z_a'' &= \frac{1}{0.11639jx} + \frac{33.075jx}{6.583 - x^2} \\ z_b'' &= \frac{6.2893jx}{6.583 - x^2} \end{aligned}\right\}. \tag{926a}$$

The common value of $j\,1.58$ which occurs at $x = 1.2629$ is not useful, nor is it necessary to consider this since the functions have a common zero at infinity which will remain after removing the pole. This reactance is

$$z_2 = \frac{6.2893jx}{6.583 - x^2} = \frac{1}{0.15901jx + \dfrac{1}{0.95534jx}}. \tag{925a}$$

There remains

$$\left.\begin{aligned} z_a''' &= \frac{1}{0.11639jx} + \frac{1}{0.03733jx + \dfrac{1}{4.069jx}} \\ z_b''' &= 0 \end{aligned}\right\}. \tag{926b}$$

With (924), (925), (925*a*), and the remainder (926*b*) we get the final

network of Fig. 173 which is equivalent to that of Fig. 172. The total number of elements is 13; the number of coils 5. The values L_0 and C_0 in the figures indicate the total inductance and capacitance (on a

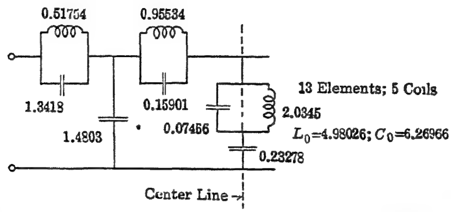


FIG. 173.—Equivalent ladder development to the one shown in Fig. 172. The parameter values are obtained from the relations (924), (925), (925a), and (926b).

$R = \omega_1 = 1$ basis), and should be considered also in estimating relative economies.

Still other equivalents may be attempted by starting with the re-

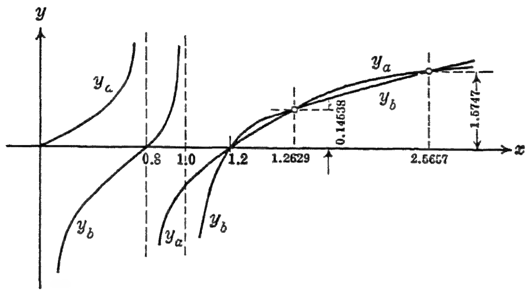


FIG. 174.—Curves showing the significant features in the behavior of the susceptance functions (927) which are the inverse of the reactances plotted in Fig. 164.

ciprocals of (891a), i.e., with the corresponding susceptance functions. Expanding these in partial fractions we have

$$\left.\begin{aligned} y_a &= \frac{0.56738jx}{0.64 - x^2} + 0.70922jx \\ y_b &= \frac{1}{1.4703jx} + \frac{0.1169jx}{1 - x^2} + 0.73801jx \end{aligned}\right\}. \tag{927}$$

These are sketched in Fig. 174. Although they have a common pole at infinity, this cannot be entirely removed from y_a since the remaining functions will not have a common zero. Such a common zero may be created by taking out either of the common values

$$\left.\begin{aligned} y_a = y_b &= j0.14538 \quad \text{at} \quad x = 1.2629 \\ y_a = y_b &= j1.5747 \quad \text{at} \quad x = 2.5657 \end{aligned}\right\}. \tag{928}$$

Taking out the smaller which corresponds to a susceptance

$$y_1 = \frac{0.14538jx}{1.2629} = 0.11512jx, \tag{929}$$

leaves

$$\left.\begin{aligned} y_a' &= \frac{0.56738jx}{0.64 - x^2} + 0.5941jx = \frac{0.5941jx(1.595 - x^2)}{0.64 - x^2} \\ y_b' &= \frac{1}{1.4703jx} + \frac{0.1169jx}{1 - x^2} + 0.62289jx \\ &= \frac{(0.68451 - x^2)(1.595 - x^2)}{1.6054jx(1 - x^2)} \end{aligned}\right\}, \tag{928a}$$

or

$$\left.\begin{aligned} z_a' &= \frac{1}{1.4806jx} + \frac{1.0078jx}{1.595 - x^2} \\ z_b' &= \frac{0.55627jx}{0.68451 - x^2} + \frac{1.0491jx}{1.595 - x^2} \end{aligned}\right\}. \tag{928b}$$

These have a common zero at infinity in addition to the pole at $x = 1.2629$. Hence we can take out

$$z_1 = \frac{1.0078jx}{1.595 - x^2} = \cfrac{1}{0.99226jx + \cfrac{1}{0.63185jx}}, \tag{929a}$$

leaving the reactances

$$\left.\begin{aligned} z_a'' &= \frac{1}{1.4806jx} \\ z_b'' &= \frac{0.55627jx}{0.68451 - x^2} + \frac{0.0413jx}{1.595 - x^2} = \frac{0.59757jx(1.5321 - x^2)}{(0.68451 - x^2)(1.595 - x^2)} \end{aligned}\right\}, \tag{928c}$$

or the susceptances

$$\left.\begin{aligned} y_a'' &= 1.4806jx \\ y_b'' &= \frac{1}{0.83855jx} + \frac{0.05823jx}{1.5321 - x^2} + 1.6734jx \end{aligned}\right\}. \tag{928d}$$

Here we take out

$$y_2 = 1.4806jx, \tag{929b}$$

leaving

$$\left.\begin{aligned} y_a''' &= 0 \\ y_b''' &= \frac{1}{0.83855jx} + \cfrac{1}{17.1722jx + \cfrac{1}{0.03801jx}} + 0.1928jx \end{aligned}\right\}. \tag{928e}$$

The final network is shown in Fig. 175. It involves a total of 12 elements of which 4 are coils.

Finally we may attempt a fourth development by again starting with the susceptances (927) but taking out the larger of the common values (928).

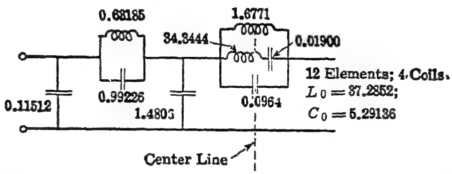


FIG. 175.—Equivalent ladder development to the ones shown in Figs. 172 and 173. The parameter values are obtained from the relations (929), (929*a*), (929*b*), and (928*e*).

This is a susceptance

$$y_1 = \frac{1.5747jx}{2.5657} = 0.61375jx. \tag{930}$$

After inverting the remainder and expanding in partial fractions we have

$$\left.\begin{aligned} z_a' &= \frac{1}{0.982jx} + \frac{9.4562jx}{6.583 - x^2} \\ z_b' &= \frac{0.23618jx}{0.8312 - x^2} + \frac{7.8115jx}{6.583 - x^2} \end{aligned}\right\}. \tag{931}$$

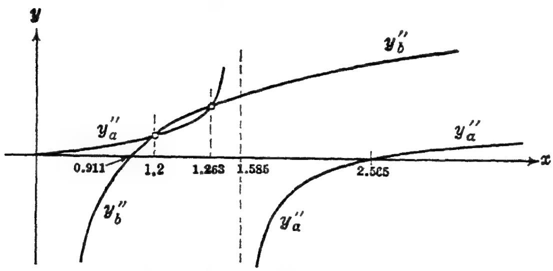


FIG. 176.—Curves showing the significant features in the behavior of the susceptance functions (931*a*). The common values at $x = 1.2$ or $x = 1.263$ are not useful in providing for a continuation in the ladder development.

These have a common zero at infinity in addition to the common pole. Hence we take out

$$z_1 = \frac{7.8115jx}{6.583 - x^2} = \frac{1}{0.12802jx + \dfrac{1}{1.1866jx}}. \tag{930a}$$

Again inverting the remainder and expanding we find

$$\left.\begin{aligned} y_a'' &= \frac{1.5267jx}{2.5173 - x^2} + 0.37551jx \\ y_b'' &= \frac{1}{0.28414jx} + 4.2341jx \end{aligned}\right\}, \tag{931a}$$

which are sketched in Fig. 176. If we remove the pole at infinity of y_a'' entirely, the remaining functions will have no common zero, and we shall not be able to continue. If we attempt to remove a susceptance which will create a common zero at one of the points $y_a'' = y_b''$, however, the remainder of y_a'' will evidently have a negative slope over

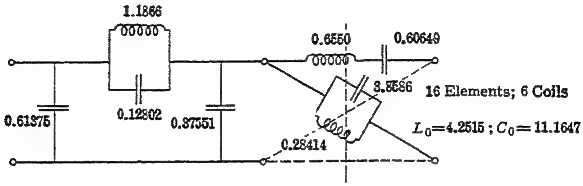


FIG. 177.—Equivalent development, not completely in the ladder form, to the ones shown in Figs. 172, 173, and 175. The parameter values are obtained from the relations (930), (930a), (930b), and (931b).

part of the region from the origin to that zero. The common values, therefore, are of no use. We can carry the development only one step farther by taking out

$$y_2 = 0.37551jx, \tag{930b}$$

leaving

$$\left.\begin{aligned} y_a''' &= \cfrac{1}{0.6550jx + \cfrac{1}{0.60649jx}} \\ y_b''' &= \frac{1}{0.28414jx} + 3.8586jx \end{aligned}\right\}. \tag{931b}$$

This undeveloped remainder must be left in the lattice form.[1] The resulting network is shown in Fig. 177.

[1] Since this portion of the manuscript was written, Brune has pointed out that the common values of y_a'' and y_b'' may be utilized to create common zeros even though this process leads to a negative inductance in the subsequent series arm of the development since such a negative inductance element may be absorbed by means of mutual coupling with the inductance in the preceding series arm after the fashion discussed by Brune in his paper "Synthesis of a Finite Two-terminal Network Whose Driving-point Impedance is a Prescribed Function of Frequency," Jl. Math. & Phys., Oct. 1931, pp. 191–236. A similar result, however, would be obtained by a conversion of the central lattice of Fig. 177 into the equivalent form shown for the general case in Fig. 181. This conversion is discussed below, and in the simplest case leads to a T-section with mutual inductive coupling between its series arms bridged by a single condenser.

Still other ladder forms may be obtained by going out from the reciprocals of z_a and z_b, i.e., replacing the original reactances by their inverse values. It will be recalled from section 4 that this does not affect the propagation function (except for an additive π in the phase)

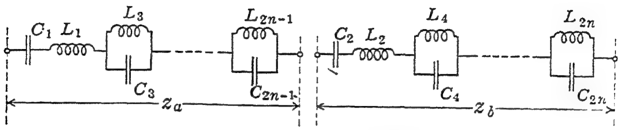


FIG. 178.—Component reactances of a lattice-designed filter in the form of anti-resonant components in series.

and merely replaces z_0 by its inverse. The process is similar to using a Π in place of a T-structure in the uniform ladder. The networks which result from this process are simply given by reciprocating those already obtained with respect to unity. This method replaces shunt by series branches, and vice versa, the reactances in these branches being replaced by their inverses. Thus inductance values become capacitance values (on a $R = \omega_1 = 1$ basis), and vice versa. Since the actual inductances and capacitances are these values multiplied by R/ω_1 and $1/R\omega_1$, respectively, it is seen that for a given case the reciprocated structures present different relative inductance and capacitance values. For example, if a large R and small ω_1 leads to excessively large inductances and small capacitances in the original network, this situation may be improved in the reciprocal network.

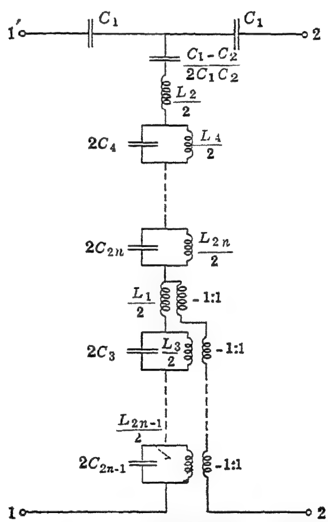


FIG. 179.—Equivalent lattice network, suggested by Cauer, for the realization of the filter whose component reactances have the form shown in Fig. 178.

Besides the ladder developments there are other networks, equivalent to the lattice, which may be suitable for the practical realization of the filter. These were discussed in section 7 of Chapter IV and are illustrated in Figs. 49 and 50 on page 161. One practical objection to these equivalents is due to the fact that they require ideal transformers which can be only approximately realized. In some cases this objection may be overcome partly or wholly. Another objection, however, is that which was

raised with regard to the lattice itself, namely, that high values of attenuation are difficult to obtain with a single section. Although not in the form of a Wheatstone bridge, these equivalents are subject to the same difficulties.

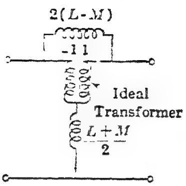


FIG. 180 — Equivalent circuit of an actual transformer (neglecting loss) according to the lattice equivalent of Fig. 50, page 161

If the lattice reactances are in the form of anti-resonant components in series, as shown in Fig. 178, then, as suggested by Cauer, a modification of the network of Fig. 49(a) results in the equivalent given by Fig. 179. Although this result does not involve an ideal transformer, it does call for close-coupled coils, i.e., pairs of identical coils with zero leakage between them. At low frequencies where iron may be used, this can be fairly well realized.

A somewhat more promising equivalent, based upon the network of Fig. 50 (page 161), is suggested by Bode. The ideal transformer difficulty is here treated in the following manner. If we consider an actual 1:1 ratio transformer, then the open-circuit input and transfer impedances for neglected losses are given by

$$z_{11} = j\omega L = z_{22};\ z_{12} = j\omega M, \tag{932}$$

where L is the inductance of its identical windings and M the mutual inductance. The equivalent lattice reactances are, therefore,

$$\left.\begin{aligned} Z_{at} &= z_{11} - z_{12} = j\omega(L - M) \\ Z_{bt} &= z_{11} + z_{12} = j\omega(L + M) \end{aligned}\right\}. \tag{933}$$

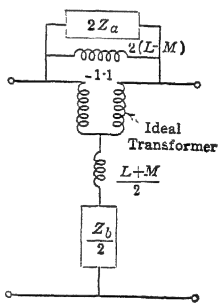


FIG. 181.—Resulting network, suggested by Bode, for the realization of a lattice-designed filter according to Fig. 50 in which the ideal transformer is replaced by an actual one whose schematic is represented as in Fig. 180.

Using these in the arrangement of Fig. 50 results in the equivalent transformer circuit shown in Fig. 180. It follows, therefore, that if we make use of the lattice equivalent shown in Fig. 50 as a means for physically realizing a given filter design, the actual circuit arrangement, except for losses, is that shown in Fig. 181 in which the actual transformer equivalent of Fig. 180 replaces the ideal. If the reactance Z_a (which may be in the form of resonant components in parallel) has a parallel inductance *less than or equal to* $(L - M)$, and if the reactance Z_b (which may be in the form of anti-resonant components in series) has a series inductance *greater than or equal to* $(L + M)$, then the transformer leakage and winding inductances can be absorbed by the filter reactances, thus

overcoming the difficulty of realizing the ideal transformer completely. It should be noted that the numerical sign of M depends upon the relative winding sense of the actual transformer and, therefore, is controllable. Thus if the available inductances in Z_a and Z_b make it more advantageous for $(L - M)$ to be less than $(L + M)$, the transformer may be connected so as to make M positive; if the reverse is the case, M can be made negative. This equivalent has the further advantage that, if the distributed capacitance between the primary and secondary windings of the transformer is appreciable, this will appear (to a first approximation) as a capacitance in parallel with the reactance $2Z_a$, and also may be absorbed by this reactance if an equally large or larger parallel capacitance is contained in it.

When a filter is designed with the Bode method of parameter determination in y_0 alone or in both the functions y_0 and z_0, the ladder development cannot be carried to completion. It is not difficult to appreciate why this should be so, in view of the fact that the points at which $z_a = z_b$ or $y_a = y_b$ (common values) serve as means to produce common zeros and hence common poles in the subsequent functions after inversion. Since the Bode method does not lead to real roots of the equation $y_0 = 1$ (equivalent to $z_a = z_b$), this process is not available. In such a case we may be left with an undeveloped portion whose minimum attenuation values are higher than can be realized by a single lattice structure without practical difficulty. An attempt may then be made to break this remainder up into a cascade of simpler lattices whose individual minimum attenuation values are low enough for practical realization. A difficulty presents itself here because this remainder by itself may not be a filter at all in the sense that a logical process of decomposition can successfully be applied to it.[1] In this case, the method of breaking the lattice into simpler components, however, can be applied to the original network so that, in the end, the difficulty involves merely a question of economy.

The method of decomposition referred to is based upon the fact that a complicated lattice can be replaced by a cascade of lattices provided their characteristic impedances are all equal to that of the original and the sum of their propagation functions equals that of the original. If the characteristic impedance function of the original is given by z_0, then all individual lattices are determined for this same function. If the index function of the original is given by y_0, and if we denote the

[1] In a general sense the remainder, of course, is still a filter, but perhaps of an entirely different class; i.e., it may have an altogether different allocation of transmission and attenuation bands from that of the structure as a whole.

index functions of the individual lattices by $y_1, y_2, y_3, \ldots$, then we must evidently have

$$\frac{y_0 + 1}{y_0 - 1} = \frac{y_1 + 1}{y_1 - 1} \cdot \frac{y_2 + 1}{y_2 - 1} \cdot \frac{y_3 + 1}{y_3 - 1} \cdots \tag{934}$$

Taking the low-pass case as an example, the simplest index function for an individual lattice according to Table II is given by[1]

$$y_i = \frac{\sqrt{\omega^2 - \omega_1^2}}{m_i \omega}. \tag{935}$$

This corresponds to an attenuation function having one infinite maximum at

$$\omega_{i\infty} = \frac{\omega_1}{\sqrt{1 - m_i^2}}, \tag{936}$$

and a minimum attenuation value of

$$(\gamma_{1i})_{\min} = \ln\left(\frac{1 + m_i}{1 - m_i}\right) \tag{937}$$

occurring at $\omega = \infty$.[2] The index functions y_i for the individual lattices, therefore, are determined completely by specifying the cut-off frequency common to all, and the frequencies $\omega_{i\infty}$ (which determine the m_i's) at which the infinite maxima in the attenuation should occur. Since the relation (934) uniquely determines y_0 in terms of $y_1, y_2, \ldots$, it follows that *the index function (propagation function) of a filter is uniquely determined by the specification of the cut-off frequency (or frequencies) plus the frequencies corresponding to infinite attenuation.*

This fact, which was first pointed out by Bode, is useful in filter design in many ways. In our present case, for example, it immediately determines the number of individual lattices into which a more complicated one is to be split, and also the index function for each. For a given characteristic impedance function z_0, the individual lattices are thus completely determined.

Several additional remarks are necessary in order to clarify this procedure. Suppose an index function y_0 has been determined for a given design by any method whatsoever. In order to carry out the decomposition into component networks, the y_i's and hence the m_i's, for each

[1] The constant multiplier is here written as $1/m$ instead of H. By applying the transformations of Table V the present treatment is extended to filters of other classes.

[2] When $m_i > 1$, the attenuation reaches a finite maximum at infinity.

must be determined. This is done by first finding the roots of the equation

$$y_0 = 1, \tag{938}$$

which are the $\omega_{i\infty}$'s, and then, from the inversion of (936) which is

$$m_i = \frac{\sqrt{\omega_{i\infty}^2 - \omega_1^2}}{\omega_{i\infty}}, \tag{936a}$$

determining the m_i's. The roots of (938), however, may in general be real (which is the case if the parameters in y_0 have been determined by Cauer's method) or complex (which occurs in Bode's parameter determination). If they are complex, pure imaginary roots may also occur. When complex roots are present they, of course, must come in conjugate pairs. Points of infinite attenuation then do not occur at real frequencies.

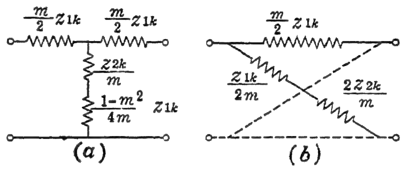


FIG. 182.—Mid-series derived Zobel-type structure (single derivation in terms of the constant-k prototype), and its lattice equivalent.

It is clear from (936a) that real values of $\omega_{i\infty}$ (greater than ω_1, of course) give rise to real values of m_i which lie between zero and unity. This corresponds to y_i-functions which are identical to those obtained by the Zobel method by the single m-derivation from the constant-k structure.[1] For a mid-series constant-k type of characteristic impedance this leads to the familiar network shown in part (a) of Fig. 182,[2] the lattice equivalent of which is given in part (b) of the same figure

[1] This is seen readily if we recall that for the constant-k filter the index function according to (749a), p. 328, is

$$y_k = \sqrt{1 - x_k^{-2}},$$

and that for the single m-derived type it is given by

$$y_{km} = \frac{y_k}{m} = \frac{\sqrt{x_k^2 - 1}}{m x_k},$$

as expressed by (791), p. 345. Since for the low-pass filter $x_k = \omega/\omega_1$, this checks with the form of (935). The extension to filters of other classes is made either by substituting the proper relation for x_k as discussed in the previous chapter (which restricts us to the even functions, however), or by making use of the transformations in Table V as already pointed out.

[2] This is represented in terms of the constant-k component reactances z_{1k} and z_{2k}, thus making more evident the application of the following results to other filter classes.

(the dotted lines indicate corresponding reactances to those in the other arms). From the standpoint of the propagation function alone, a Cauer design, therefore, is immediately realized by a cascade of Zobel-type sections corresponding to the various points of infinite attenuation as determined by (882) with the proper frequency transformation in Table V. When a simple characteristic impedance function is sufficient, this gives the realization of a Cauer design directly in the ladder form.

In the case of pure imaginary values of $\omega_{i\infty}$, (936a) shows that the corresponding m_i-values are real but greater than unity. It will be recalled that such values of m_i give rise to a more linear or even convex phase characteristic. The structure, although not generally realizable in the form of a T or Π, is still realizable as a lattice.

When the values of $\omega_{i\infty}$ are complex, the corresponding values of m_i are also complex, which means that the individual lattices can no longer be built. However, since complex values always occur in conjugate pairs, they determine pairs of lattices with conjugate values of m_i. Such a cascaded pair of lattices can always be combined into a single realizable lattice. This may be seen by applying the principles discussed in section 5 of Chapter IV for finding the resultant of two cascaded structures. Bearing in mind that the two cascaded lattices have the same characteristic impedance, the resultant lattice corresponding to a pair of conjugate m-values (denoted by m and $\overline{m}$) is that shown in Fig. 183. The factors involving the m's are seen to be real.

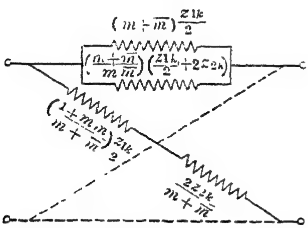


FIG. 183.—Resultant lattice corresponding to a cascade of two lattices, of the form shown in Fig. 182(b), for a pair of conjugate complex m-values.

It is thus clear that any given design can always be realized in the form of cascaded structures each of which accounts for one of the roots (or a pair of conjugate roots) of the equation $y_0 = 1$. The sections corresponding to real roots are always realizable as ladder structures without restriction on the characteristic impedance function.[1] When all roots are real, a corresponding ladder development is thus given directly. Since, in this process, each section has the same characteristic

[1] When the characteristic impedance function is of higher order than that of the constant-k type (which corresponds to the function $F_2(x)$), then the component index function y_i together with the specified z_0-function is used to form a corresponding pair of lattice reactances from which a ladder development is found according to the method already discussed. It is only in the case where z_0 corresponds to $F_2(x)$ that the component structure is in general recognizable directly as a Zobel type.

impedances, the method evidently leads to poor economy except in those cases where a simple z_0-function is required.

The possibility of having sections with m-values larger than unity or with complex m-values, shows a considerable increase in the variety of results obtainable as compared with the Zobel structures. In this connection it is also useful to recognize that the lattices which are obtained for the realization of sections with such m-values sometimes may be converted into equivalent bridged-T structures. This conversion is most easily seen with the help of *Bartlett's bisection theorem*[1] which states

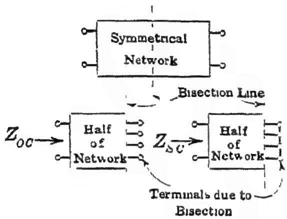


FIG. 184.—Schematic representation of the open and short-circuit impedances (939) for a symmetrical network characterised by Z_0 and γ, according to Bartlett's bisection theorem.

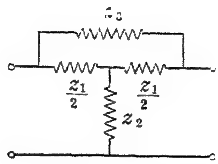


FIG. 185.—Possible bridged-T equivalent for the lattice-designed filter section of Fig. 183 according to the relations (941) and (941a)

that, if Z_0 and γ are the characteristic impedance and propagation functions of a symmetrical network, then

$$\left.\begin{aligned} Z_{oc} &= Z_0 \coth \frac{\gamma}{2} \\ Z_{sc} &= Z_0 \tanh \frac{\gamma}{2} \end{aligned}\right\} \tag{939}$$

are the open- and short-circuit impedances of half the network, as illustrated in Fig. 184. Equation (432), page 180, however, shows that $Z_{sc} = Z_a$ and $Z_{oc} = Z_b$ are the component impedances of a symmetrical lattice defined by Z_0 and γ. Hence any network whose open- and short-circuit impedances after bisection equal Z_b and Z_a, respectively, is equivalent to the symmetrical lattice given by these impedances.

[1] A. C. Bartlett, "The Theory of Electrical Artificial Lines and Filters," John Wiley & Sons, 1930, p. 28. A simpler and more enlightening proof is given by O. Brune, "Note on Bartlett's Bisection Theorem," Phil. Mag., Ser. 7, Vol. 14, p. 806, November, 1932.

For the bridged-T of Fig. 185 the equivalent lattice impedances are, therefore,

$$\left.\begin{aligned} Z_a &= \frac{1}{\dfrac{1}{z_1/2} + \dfrac{1}{z_3/2}} \\ Z_b &= \frac{z_1}{2} + 2z_2 \end{aligned}\right\}. \tag{940}$$

Reference to Fig. 183 shows that if we let

$$\left.\begin{aligned} \frac{z_3}{2} &= \frac{m + \overline{m}}{m\overline{m}}\left(\frac{z_{1k}}{2} + 2z_{2k}\right) \\ \frac{z_1}{2} &= (m + \overline{m})\,\frac{z_{1k}}{2} \end{aligned}\right\}, \tag{941}$$

then it follows that

$$2z_2 = \frac{2z_{2k}}{m + \overline{m}} + \left\{\frac{1 + m\overline{m}}{m + \overline{m}} - (m + \overline{m})\right\}\frac{z_{1k}}{2}. \tag{941a}$$

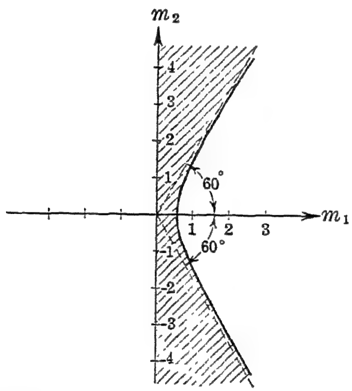


FIG. 186.—Complex m-plane in which the shaded area represents the region of m-values for which the bridged-T equivalent of Fig. 185 is physically realizable.

This impedance is realizable so long as

$$\frac{1 + m\overline{m}}{m + \overline{m}} \geq m + \overline{m}.$$

If we write $m = m_1 + jm_2$, then this condition reduces to

$$m_1 \leq \sqrt{\frac{1 + m_2^2}{3}}. \tag{942}$$

In the complex m-plane this represents the shaded area bounded by the hyperbola as illustrated in Fig. 186. Any m-vector within this

region, therefore, leads to a physically realizable bridged-T equivalent to the lattice of Fig. 183, with components given by (941) and (941a). Recognizing that in the lattice z_{1k} occurs *six* times and z_{2k} *four* times, while in the bridged-T z_{1k} occurs *four* times and z_{2k} *two* times, makes evident the saving in elements which may be had by this transformation.

11. Dissymmetrical filters. In practice it frequently occurs that a filter is to work out of and into different resistances. Although the process of designing dissymmetrical networks for this purpose is still in a very unsatisfactory state, certain guiding principles may be worth mentioning here. A useful theorem in this connection is due to Bode. This states that, when the propagation function and image impedance at one end of a reactive network are known, the image impedance at the other end is determined to within a constant multiplier. The proof of this is readily given. By means of the relations (291) and (292), pages 137 and 138, the following identity is seen to be correct

$$\left(\frac{z_{22}}{z_{11}}\right) \cdot \left(\frac{y_{11}z_{11} - 1}{y_{11}z_{11}}\right) = \left(\frac{z_{12}}{z_{11}}\right)^2. \tag{943}$$

Since z_{11} and z_{12} are pure imaginaries, the ratio on the right-hand side is real and positive at all real frequencies. This means that the factors on the left (which are also real) must either both be positive or both negative; i.e., they must change sign at the same frequencies. When y_{11} and z_{11} are given, this fact, therefore, enables us to determine the critical frequencies (poles and zeros) of z_{22}. With these known, z_{22} is determined to within a constant multiplier; i.e., the reactive network is specified completely by the two functions y_{11} and z_{11} (which refer to end 1) except for a constant multiplier.

Since by (413) and (414), page 175,

$$\frac{z_{22}}{z_{11}} = \frac{Z_{I2}}{Z_{I1}}; \quad \frac{y_{11}z_{11} - 1}{y_{11}z_{11}} = 1 - \tanh^2\theta, \tag{944}$$

the condition resulting from (943) may be written alternatively as

$$\left(\frac{Z_{I2}}{Z_{I1}}\right) \cdot (1 - \tanh^2\theta) \geq 0. \tag{943a}$$

The theorem, therefore, may be alternatively stated by saying that, if the critical frequencies (controlling parameters) of Z_{I1} and the function θ are given, the critical frequencies of Z_{I2} are determined thereby.

In a transmission range, y_{11} and z_{11} have the same sign, and the second factor in (943) remains positive. Hence the poles and zeros of z_{22} coincide with those of z_{11} and y_{11}.

In an attenuation range, y_{11} and z_{11} have opposite signs ($y_{11}z_{11}$ is real and positive) so that the second factor in (943) changes sign only at

roots of odd multiplicity of the equation $y_{11}z_{11} = 1$. There z_{22} is critical except when these points coincide with critical frequencies of z_{11} (which make the first factor in (943) change sign also). z_{22} is also critical at the remaining critical frequencies of z_{11}.

At the boundaries between these regions (which are due to critical frequencies of y_{11} not contained in z_{11}, or vice versa), the second factor in (943) evidently changes sign only when such points are due to zeros in either y_{11} or z_{11}. Hence z_{22} is critical only if the cut-off is due either to a zero of y_{11} or a pole of z_{11}.

If we attempt to obtain a dissymmetrical filter by simply bisecting a symmetrical one, these results give us some idea as to what we may expect in the way of an image impedance at the point of bisection. If we regard y_{11} and z_{11} as referring to the original end of half of a bisected equivalent lattice structure whose reactances are Z_a and Z_b, then

$$\frac{1}{y_{11}} = Z_a; \; z_{11} = Z_b,$$

and

$$y_{11}z_{11} = y_0^2.$$

When the parameters in y_0 are determined on the Cauer basis, the equation $y_0 = 1$ has simple roots at real frequencies (points of infinite attenuation). z_{22}, therefore, can never have the same distribution (although it may have the same number) of critical frequencies in the attenuation range as z_{11}. Since the number and position of the latter are selected so as to give the image impedance Z_{I1} a favorable form in the transmission range, it follows that the image impedance Z_{I2} at the bisected end may be expected to have almost any form at all. Unless the design can be suitably planned, this process, therefore, will not be successful.

When the Bode method of parameter determination is used for y_0, the equation $y_0 = 1$ has no roots at real finite frequencies, and hence Z_{I2} has the same parameters as Z_{I1}. The bisected filter then shows the same terminal conditions at both ends. However, since the network remains partly in the lattice form, a bisection is not possible.

The fact that Z_{I2} is thus determined except for a multiplying factor means that the ratio of impedance levels at the two ends of the bisected filter is not determined by the reactances Z_a and Z_b but depends upon the form of the network which is used for its realization. As we have seen from the discussion in the previous section, this may be varied in a number of ways. At the present state of this process we are unable to give any guiding principles whereby the ratio of impedance levels may be predicted or controlled.

One helpful point of view in this connection is offered by the following transformation. Suppose a certain portion of the ladder development has the form of network (*a*) in Fig. 187. This is equivalent to the network (*b*) which includes the ideal transformer if

$$\left.\begin{aligned} z_1 + z_2 &= z_A + z_C \\ z_2 &= a z_C \\ z_2 &= a^2(z_B + z_C) \end{aligned}\right\}, \tag{945}$$

from which

$$\left.\begin{aligned} z_A &= z_1 - z_2\left(\frac{1-a}{a}\right) \\ z_B &= z_2\left(\frac{1-a}{a^2}\right) \\ z_C &= \frac{z_2}{a} \end{aligned}\right\}. \tag{945a}$$

If z_1 is such a reactance that we may write

$$z_1 = z_1' + b z_2 \tag{946}$$

with z_1' physical and b real and positive, then we see that the reactances z_A, z_B, z_C are physical provided

$$1 \geq a \geq \frac{1}{b+1}. \tag{946a}$$

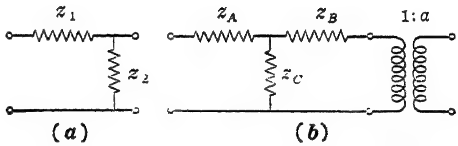


FIG. 187.—Network equivalence, according to the relations (945) to (946*a*), which may be useful in obtaining a change in impedance level.

For any value of b greater than zero, the ratio a of the ideal transformer can be less than unity, which means that the network (*b*) without the ideal transformer is equivalent to (*a*) with a corresponding change in the ratio of impedance levels at its two ends. In the present example the omission of the ideal transformer steps the ratio up from left to right or down from right to left. Whether a step-up or step-down in a given direction is possible, evidently depends upon the orientation of the network (*a*) in the given ladder structure.

It is evident, of course, that this method of gaining a change in the ratio of impedance levels is not restricted to the bisected network. It should be understood also that the removal of the ideal transformer from its normal position to either end of the complete structure necessitates a corresponding change in impedance level of all elements in that portion of the structure between the normal position and the end in question. For example, if a network N forms the balance of the

structure from the right-hand terminals of network (a) of Fig. 187 to the right-hand terminals of the complete structure, then the removal of the ideal transformer to these terminals involves multiplying all inductances in N by $1/a^2$ and all capacitances by a^2. Dropping the ideal transformer then amounts to changing the impedance level at that end by the factor $1/a^2$. The success of the method depends upon having the reactances z_1 and z_2 satisfy (946), and in that event the largest gain in ratio is restricted to the condition (946a). In an extensive structure it may, of course, be possible to make similar transformations at several points.[1] The process does, incidentally, nicely illustrate the fact that the impedance level at one end of a network is not determined by the image propagation function and the image impedance at the other end.

12. Reflection losses. As in the conventional filter, the net loss is given by the sum of the attenuation function and those losses due to mismatching at the terminals. These are substantially represented by the term designated as the reflection loss, as pointed out in section 10 of Chapter IX. This loss, it will be recalled, has infinite maxima at the zeros and poles of the characteristic impedance function, and negative minima equal to ln2 napiers at those frequencies where the magnitude of the characteristic impedance equals the terminal resistance R. For the constant-k filter with m-derived terminal networks, we saw that the infinite maxima in the reflection loss coincide with those in the attenuation function, and that the negative minima occur in the vicinity of the minima of the attenuation function. This is not true of the lattice filter since the attenuation and characteristic impedance functions are entirely independent. Here no general conclusions can be drawn for the reason that each case presents a different situation with regard to the relative distribution of infinite attenuation maxima and the zeros and poles of the characteristic impedance function.[2]

[1] Transformations similar to these may also be used for the purpose of arriving at other, more reasonable values of inductance and capacitance than those in a given physical realization. For an application of this method see "Electric Wave Filters for High Frequencies," C. L. Frederick, Electronics, March, 1934, p. 84.

[2] When the parameters in y_0 are determined on the Bode basis, the attenuation function shows no infinite maxima. In this case the net loss shows such maxima at those points where the reflection loss is infinite. In the Cauer filter the reflection loss will produce, in general, additional infinite maxima. In a low-pass filter, for example, the reflection loss causes the net attenuation to become infinite at $\omega = \infty$. The negative minima, on the other hand, although they in general do not coincide with the minima of the attenuation function, will cause a modification in the resulting loss function of the filter. This fact must be accounted for in a carefully considered design.

13. The effect of incidental dissipation.[1] Although the determination of the exact effect which the losses in the coils and condensers have upon the filter behavior is a long and tedious process, certain approximate relations can be deduced which materially aid in predicting this effect so that it may be taken into account, to a certain degree, in the design. In this respect it may be stated at the outset that incidental dissipation affects the performance of a filter primarily within its transmission ranges, to the extent that it brings about slight departures from the predicted attenuation and phase shift.

To begin with we shall assume that all the coils and condensers in the network have the same resistance-to-inductance and leakage conductance-to-capacitance ratios, respectively, but not necessarily that the R/L ratio equals the G/C ratio. This is sufficiently close to practical conditions to be a good working basis. Such a uniformly dissipative network was discussed in section 2 of Chapter VII in connection with the lattice-type artificial line (see pages 268–269). As a result of this discussion it follows that if $Z(\lambda)$ is the impedance function of a non-dissipative network, then $\mu \cdot Z(\lambda')$ is the impedance of the corresponding uniformly dissipative network where

$$\left.\begin{aligned} \mu &= \sqrt{\frac{\lambda + 2\alpha}{\lambda + 2\beta}} \\ \lambda' &= \sqrt{(\lambda + 2\alpha)(\lambda + 2\beta)} \end{aligned}\right\}, \tag{947}$$

and

$$2\alpha = \frac{R}{L}, \; 2\beta = \frac{G}{C} \tag{948}$$

are the uniform ratios.

The relations (947) may be rewritten as

$$\left.\begin{aligned} \mu &= \left\{1 + \frac{2(\alpha - \beta)}{\lambda + 2\beta}\right\}^{\frac{1}{2}} \\ \lambda' &= (\alpha + \beta + \lambda)\left\{1 - \left(\frac{\alpha - \beta}{\alpha + \beta + \lambda}\right)^2\right\}^{\frac{1}{2}} \end{aligned}\right\}, \tag{947a}$$

[1] The fundamental idea of the method described in this section is due to H. F. Mayer and is given in his paper "Ueber die Daempfung von Siebketten im Durchlaessigkeitsbereich," E. N. T. Vol. 2, No. 10, 1925, pp. 335–338. In a subsequent paper "Daempfung und Winkelmass von Vierpolen mit geringen Verlusten," T. F. T. July 1932, pp. 179–187, A. Feige and F. Holzapfel give consideration particularly to the extent of the region of convergence of the expansions involved in this process. Neither paper describes, however, the fundamental point of departure, which is based upon the relations between the impedance functions of the uniformly dissipative and the non-dissipative networks.

or expanding by the binomial theorem we have

$$\left.\begin{aligned} \mu &= 1 + \frac{\alpha - \beta}{\lambda + 2\beta} - \cdots \\ \lambda' &= (\alpha + \beta + \lambda)\left\{1 - \tfrac{1}{2}\left(\frac{\alpha - \beta}{\alpha + \beta + \lambda}\right)^2 - \cdots\right\} \end{aligned}\right\}. \tag{947b}$$

These may be replaced by their first terms provided the corresponding errors, as indicated approximately by the magnitudes of the second terms in the expansions, are negligibly small. Thus for real frequencies ($\lambda = j\omega$) we may write

$$\left.\begin{aligned} \mu &= 1 \\ \lambda' &= \alpha + \beta + j\omega \end{aligned}\right\}, \tag{949}$$

provided the following approximate criteria are satisfied, respectively,

$$\left.\begin{aligned} \left(\frac{R}{2L\omega} - \frac{G}{2C\omega}\right) &<< 1 \\ \tfrac{1}{2}\left(\frac{R}{2L\omega} - \frac{G}{2C\omega}\right)^2 &<< 1 \end{aligned}\right\}. \tag{949a}$$

In the special case $\alpha = \beta$ the relations (949) are exact at all frequencies. Otherwise the criteria (949a) determine the approximate errors involved. The errors evidently decrease with increasing frequency. When the approximations (949) are justified, we see that the dissipative impedance is obtained from the non-dissipative (reactance function) by simply replacing $j\omega$ by $(\alpha + \beta) + j\omega$. Except for the vicinity of the origin, where the criteria (949a) are not satisfied, this result shows that the impedance of a uniformly dissipative network is found for real frequencies by evaluating the corresponding reactance function along a straight line $(\alpha + \beta)$ units to the right of the imaginary axis in the λ-plane.

In the case of the propagation function, which is of primary interest to us at present, this may be done when the second criterion (949a) alone is satisfied. This is true because the propagation function depends only upon the *ratio* of two impedances so that the factor μ does not need to be considered.

If, for the non-dissipative filter, we denote the propagation function by

$$P_0 = A_0 + jB_0, \tag{950}$$

and for the corresponding uniformly dissipative network by

$$P = A + jB, \tag{950a}$$

then A and B (the dissipative attenuation and phase functions) may be

expressed in terms of the non-dissipative functions A_0 and B_0 by means of the following series expansions

$$\left.\begin{aligned} A &= A_0 + \frac{\partial A_0}{\partial \gamma}\cdot(\alpha+\beta) + \frac{1}{2}\frac{\partial^2 A_0}{\partial \gamma^2}\cdot(\alpha+\beta)^2 + \cdots \\ B &= B_0 + \frac{\partial B_0}{\partial \gamma}\cdot(\alpha+\beta) + \frac{1}{2}\frac{\partial^2 B_0}{\partial \gamma^2}\cdot(\alpha+\beta)^2 + \cdots \end{aligned}\right\}, \tag{951}$$

where γ is the real part of λ; i.e., λ is written as $\gamma + j\omega$. These expansions converge within regions where A and B are regular and continuous. Since we are interested only in their behavior in the transmission range, this condition is evidently met.

In view of the fact that the validity of these expansions is based upon the approximation discussed above, it would be improper to use them for the evaluation of A and B beyond a consistent approximation. For practical reasons this is quite sufficient. If we restrict the relations (951) to their first two terms, and make use of the condition equations involving the real and imaginary parts of a function of a complex variable,[1]

$$\left.\begin{aligned} \frac{\partial A_0}{\partial \gamma} &= \frac{\partial B_0}{\partial \omega} \\ \frac{\partial B_0}{\partial \gamma} &= -\frac{\partial A_0}{\partial \omega} \end{aligned}\right\}, \tag{952}$$

and note that in the transmission range $A_0 = 0$, we get

$$\left.\begin{aligned} A &\cong \left(\frac{R}{2L} + \frac{G}{2C}\right)\cdot\frac{\partial B_0}{\partial \omega} \\ B &\cong B_0 \end{aligned}\right\}, \tag{951a}$$

where the units are in napiers and radians, respectively.

These results show that, to a first order correction,[2] the phase characteristic is unaffected by incidental dissipation while the attenuation in the transmission ranges is proportional to the slope of the phase characteristic, the proportionality factor depending upon the R/L and G/C ratios of the coils and condensers. Thus a constant attenuation in the pass band is obtained only with a linear phase characteristic. In a Cauer or Zobel filter, for example, the slope of the phase characteristic increases materially as the cut-off is approached. As a result

[1] These relations, which are known as the *Cauchy-Riemann condition equations*, insure that the derivative of an analytic function be independent of the direction of the differential $d\lambda = d\gamma + jd\omega$.

[2] Further examination of the expansions (951) shows that only terms of *odd* order affect A while terms of *even* order affect B. Hence the relation (951*a*) for A is strictly speaking correct to within terms of second order.

of this, the incidental attenuation shows a corresponding increase. If this incidental attenuation is to be kept within a specified minimum over a given percentage of the band, the R/L and G/C ratios must be held to minimum values easily calculable from the first relation (951a). With a more linear phase characteristic the same minimum attenuation is evidently maintained with a larger allowable factor of dissipation and a consequent saving in weight, space, and cost of the resulting filter (other factors being equal). In this respect the Bode filter shows a substantial advantage as compared with an equivalent Zobel or Cauer design in actual practice. This rather interesting way in which the linearity of the phase characteristic affects the practical realization of a filter is hardly appreciated from a casual consideration of the problem.

When the R/L and G/C ratios vary with frequency, as they do in actual cases, the above result may still be applied provided the proper values are used at each frequency for which the attenuation is calculated. When mutual inductances are present, the uniformity in the dissipation factor R/L is evidently destroyed. It is found in practice that in such cases departures from predicted behavior occur which frequently exceed what reasonably may be attributed to this cause purely on a heuristic basis. This probably may be due to the fact that, when such a transformer unit has a relatively high winding inductance and coupling factor, the leakage inductance alone affects the circuit. Although the ratio of resistance to winding inductance may be reasonably low, it will be extremely large with respect to the leakage inductance, so that the net effect is similar to a coil with an exceptionally high resistance. This may be a practical reason for avoiding mutual inductance in cases where this situation exists.

It should be noted also that for a given total phase shift throughout the transmission range of a band-pass filter (the least this can be is π radians), the average slope varies inversely as the band width. Hence for the same R/L and G/C ratios a narrow band filter will have a higher incidental attenuation than a wide one. This checks with our knowledge of the behavior of the series tuned R, L, C-circuit, for example, for which the selectivity depends entirely upon the R/L ratio.

14. Concluding remarks. Bode's method of synthesis in terms of the critical frequencies of the attenuation function (cut-off frequencies plus frequencies of infinite attenuation) is of such a general nature that several additional remarks may well be made in order that the reader may more fully appreciate its value. In this connection it is of importance to recognize that the Cauer and Bode "ready made" parameter determinations should be looked upon more as guides in the

treatment of special cases requiring supplementary readjustment by means of a cut-and-try process. The "ready made" parameter determinations fit only "average cases," and by no means exhaust the possible results obtainable in filter networks generally. Special requirements will call for what we might call "tailor-made" parameter determinations for which only general guiding principles may be set down.

In order to illustrate roughly what we have in mind here, let us consider first a pair of cascaded symmetrical sections of which each has a simple index function of the "m-derived" form, thus

$$y' = \frac{y_k}{m'}; \; y'' = \frac{y_k}{m''}, \tag{953}$$

with

$$y_k = \frac{\sqrt{x^2 - 1}}{x} \tag{953a}$$

For the low-pass filter, for example, $x = \omega/\omega_1$ (where ω_1 is the cut-off frequency, which makes the variable x the same as what we called x_k in the previous chapter). The index function y_k is the same as that for the constant-k structure; i.e., the functions (953) are of the familiar Zobel m-derived type. The characteristic impedance functions are the same for both sections, but for our present purpose need not be considered in detail.

The resulting index function y_0 is found from (934) to be

$$y_0 = \frac{y'y'' + 1}{y' + y''} = \frac{(1 + m'm'')x^2 - 1}{(m' + m'')x\sqrt{x^2 - 1}} \tag{954}$$

Let us study the possibilities offered by this function in terms of the resulting attenuation and phase shift. The values m' and m'' may be real, in which case they must be positive, or conjugate complex with positive real parts. Other than this they are unrestricted. This situation becomes somewhat simpler to study if we let

$$\left.\begin{aligned} m' + m'' &= 2a \\ m'm'' &= b^2 \end{aligned}\right\} \tag{955}$$

where a and b are restricted to positive real values and

$$\left.\begin{aligned} m' &= a + \sqrt{a^2 - b^2} \\ m'' &= a - \sqrt{a^2 - b^2} \end{aligned}\right\} \tag{955a}$$

Thus if $a > b$, the m's are real and different; if $a = b$, they are real and equal; if $a < b$, they are conjugate complex.

The attenuation function depends upon the manner in which

$$\frac{1}{y_0} = \frac{2ax\sqrt{x^2 - 1}}{(1 + b^2)x^2 - 1} \tag{954a}$$

approximates unity for $1 < x < \infty$; the phase function depends upon the manner in which

$$\frac{1}{y_0} = \frac{2ajx\sqrt{1 - x^2}}{1 - (1 + b^2)x^2} \tag{954b}$$

approximates $j\tan\left(x\frac{\pi}{2}\sqrt{1 + b^2}\right)$ for the range $0 < x < 1$.

The character of the function (954*a*) depends upon whether $b < 1$ or

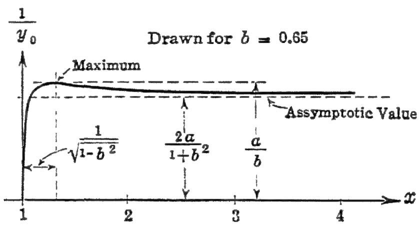


FIG. 188.—Plot showing the general character of the relation (954*a*) for the condition $b<1$.

$b \geq 1$. For $b < 1$ the result is plotted in Fig. 188. From this plot it is clear that if $\frac{2a}{1 + b^2} > 1$, only one point of infinite attenuation occurs at a real frequency; if $\frac{2a}{1 + b^2} < 1 < \frac{a}{b}$, two such points occur; and if $\frac{a}{b} < 1$, none occur. According to (955*a*) we see that the following *m*-values correspond to these possibilities

$$\frac{a}{b} > 1;\ \text{real } m\text{'s}\begin{cases} \dfrac{2a}{1 + b^2} > 1;\ m' > 1;\ m'' < 1 \\ \dfrac{2a}{1 + b^2} = 1;\ m' = 1;\ m'' < 1 \\ \dfrac{2a}{1 + b^2} < 1;\ m' < 1;\ m'' < 1 \end{cases}$$

$$\frac{a}{b} = 1;\ m' = m'' < 1 \text{ and real}$$

$$\frac{a}{b} < 1;\ \text{complex } m\text{'s},\ |m| = b < 1.$$

Thus infinite attenuation occurs at one or two real frequencies when one or both real *m*'s are less than unity, as is to be expected from our

familiarity with the Zobel theory. These occur at frequencies corresponding to

$$x = \frac{1}{\sqrt{1 - m'^2}} \quad \text{and} \quad x = \frac{1}{\sqrt{1 - m''^2}}.$$

Equal m's less than unity produce a double root of the equation $y_0 = 1$; i.e., the maximum in Fig. 188 becomes tangent to the constant *unity* ($a = b$). When this maximum is less than unity, the attenuation merely reaches a finite maximum at the corresponding frequency. This type of result is obtained only with complex m's.

If in (954a) $b \geq 1$, the maximum is suppressed, and the function has the general character shown in Fig. 189. In this case we will get one point of infinite attenuation at a real frequency only if $\frac{2a}{1 + b^2} > 1$.

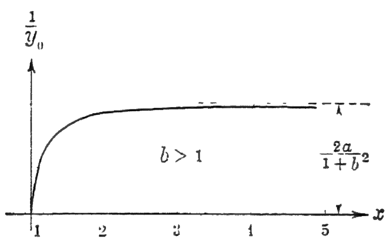


FIG. 189.—Plot showing the general character of the relation (954a) for the condition $b \geq 1$.

The various possibilities and corresponding m-values are summarized as follows:

$$\frac{a}{b} > 1;\ \text{real } m\text{'s} \begin{cases} \dfrac{2a}{1 + b^2} > 1;\ m' > 1;\ m'' < 1 \\ \dfrac{2a}{1 + b^2} = 1;\ m' > 1;\ m'' = 1 \\ \dfrac{2a}{1 + b^2} < 1;\ m' > 1;\ m'' > 1 \end{cases}$$

$$\frac{a}{b} = 1;\ \frac{2a}{1 + b^2} < 1;\ m' = m'' \geq 1 \text{ and real}$$

$$\frac{a}{b} < 1;\ \frac{2a}{1 + b^2} < 1;\ \text{complex } m\text{'s, } |m| = b \geq 1.$$

The last case, for which the attenuation rises to a finite maximum at infinity, can be obtained only with complex m-values.

In order to study the corresponding phase functions for these possibilities we may form the ratio between (954b) and the tangent function,

which is (for this case) the same as the ratio Q'/Q involving the functions discussed in section 9 above. If we write

$$x_a = \frac{1}{\sqrt{1+b^2}} \tag{956}$$

for the point at which $1/y_0$ has its pole, then this ratio becomes

$$\frac{Q'}{Q} = \frac{2ax\sqrt{1-x^2}}{1-\left(\dfrac{x}{x_a}\right)^2} \quad \frac{1}{\tan\left(\dfrac{\pi}{2}\dfrac{x}{x_a}\right)} \tag{957}$$

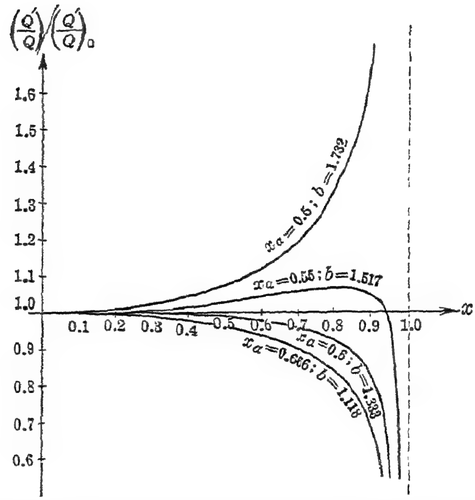


FIG. 190.—Plot of the normalized ratio (957a) for various values of the parameter x_a.

The discussion is somewhat simplified by dividing by the value for $x = 0$, which is

$$\left(\frac{Q'}{Q}\right)_{x=0} = \frac{4ax_a}{\pi}, \tag{958}$$

and thus have

$$\left(\frac{Q'}{Q}\right) \bigg/ \left(\frac{Q'}{Q}\right)_0 = \frac{\pi}{2} \cdot \frac{\dfrac{x}{x_a}\sqrt{1-x^2}}{1-\left(\dfrac{x}{x_a}\right)^2} \cdot \frac{1}{\tan\left(\dfrac{\pi}{2}\dfrac{x}{x_a}\right)}, \tag{957a}$$

which has the value unity at $x = 0$ and involves the single parameter x_a determined by the value of b according to (956). The value of a then determines the nominal value of the ratio Q'/Q while b determines its behavior versus frequency.

Fig. 190 shows a family of plots of the normalized ratio (957a) for various values of x_a. From these we see that for values of x_a somewhere between 0.55 and 0.6 a very nearly linear phase characteristic over most of the pass band is possible. The values of b for these are all larger than unity so that the corresponding behavior of $1/y_0$ in the attenuation range is given by Fig. 189. The value of a depends upon the choice of the value of Q'/Q at $x = 0$. If we choose unity as this value, then by (958)

$$a = \frac{\pi}{4x_a}, \tag{958a}$$

while if we let it be say μ, then

$$a = \frac{\mu\pi}{4x_a}. \tag{958b}$$

Defining the error by which $1/y_0$ fails to equal the tangent function by (912), i.e., by $Q'/Q = 1 + \epsilon$, we have as the error at $x = 0$

$$\epsilon = \mu - 1.$$

Choosing unity as the value of Q'/Q at $x = 0$ thus makes $\epsilon = 0$ for $x = 0$.

The deviation of the phase function from linearity is given by means of (913), and for this case becomes

$$d\gamma_2 = \epsilon \sin \pi \cdot \frac{x}{x_a}. \tag{959}$$

If a given value μ is dotted in as a horizontal line in Fig. 190, the variation of ϵ with x is easily determined graphically and (959) thus evaluated. A large variety of results are evidently possible. The reader may carry some of these out as exercises.

It is interesting to note that if we treat this case by the method of section 9 and evaluate the one parameter by means of Table VIII (this is a case for $n = 1$ and $m = 1$), the result leads to $x_a = 2/3$. Reference to Fig. 190 shows that this does not give us the most linear phase characteristic possible. This merely illustrates what we have already pointed out, namely, that Bode's parameter determination applies more specifically to large values of n. For small n it may have to be supplemented by a readjustment process. This discussion for the simple two-section case illustrates the variety of particular results obtainable if we cast aside both the Cauer and Bode "ready-made" processes, as may be found necessary in meeting the requirements of special problems.

Several comments may be made in this connection relative to addi-

tional possibilities made available by means of complex m's, as pointed out by Bode. Consider two cascaded sections with the conjugate values m and $\overline{m}$. The total attenuation is given by

$$\ln\left|\frac{y_k+m}{y_k-m}\right|+\ln\left|\frac{y_k+\overline{m}}{y_k-\overline{m}}\right|.$$

Since $m = m_1 + jm_2$ and y_k is a pure imaginary in the transmission range, it is evident that the attenuation of each section in this range is not zero, but that the attenuation of the second is the negative of that of the first. The two sections together have zero attenuation in the transmission range. In the attenuation range y_k is real so that such cancellation does not take place. An extreme case occurs when the m's are pure imaginaries. Then the ranges of transmission and attenuation for the individual sections are interchanged, and the two in cascade form an all-pass structure.

When dissipation is considered, y_k is complex; i.e., it has a real part in the transmission range and an imaginary part in the attenuation range. It then becomes clear that the attenuation values of the two sections in the transmission range are no longer exactly equal and opposite. In general the resulting incidental attenuation is small and determined approximately by the methods discussed in the previous section. A special case occurs, however, if we determine the m's so that one of them, say m, equals the complex y_k at some frequency within the nominal transmission band. This section then possesses a point of infinite attenuation at that frequency. Since at this same frequency $y_k \neq m$, the section with the conjugate m does not possess such a point of infinite attenuation, although its attenuation at all other frequencies in the nominal transmission range is nearly equal and opposite to that of the first section. The result is to introduce a rather sharp infinite peak of attenuation *within the nominal band of free transmission.* Complex m-sections are not only able to introduce sharp infinite peaks of attenuation very much closer to the theoretical cut-off than is possible with real m's, but may even produce such peaks at the cut-off or within the transmission range. This property becomes useful when a particular frequency (such as the carrier, for example) is to be eliminated. It is also useful to produce an extremely sharp cut-off. When it is attempted to do this by means of a real m, the effect of incidental attenuation so blunts the result as to defeat the purpose. The complex m-section, being a Wheatstone bridge, is able to introduce infinite attenuation (balance) in the dissipative case as well as in the non-dissipative.

A few words may be said here regarding sections with more compli-

cated characteristic impedance functions. Consider, for example, the following pair of functions defining a single section

$$y_0 = \sqrt{\frac{z_b}{z_a}} = \frac{\sqrt{1 - x^2}}{mjx};\ z_0 = \frac{H\sqrt{1 - x^2}}{\left(1 - \dfrac{x^2}{x_2^2}\right)} = \sqrt{z_b z_a}, \tag{960}$$

from which

$$z_a = \frac{Hmjx}{\left(1 - \dfrac{x^2}{x_2^2}\right)};\ z_b = \frac{H}{m} \cdot \frac{1 - x^2}{jx\left(1 - \dfrac{x^2}{x_2^2}\right)}. \tag{960a}$$

Let m be real and less than unity so that an infinite point of attenuation occurs at a real frequency, say at $x = x_\infty$. The form of the resulting ladder equivalent depends upon the location of x_∞ relative to x_2 at which z_a and z_b have a common pole. If $x_\infty < x_2$, the residue of z_b at this pole is evidently *larger* than that of z_a since $y_0 = \sqrt{z_b/z_a}$ is greater than unity beyond x_∞. Hence z_a (which consists of an anti-resonant component alone) may be removed entirely in the first step of the ladder development, leaving $z_a' = 0$ and $z_b' = z_b - z_a$ of the same form as z_b. The ladder development is completed, therefore, and is given by the network of Fig. 191 which is recognizable as a Zobel double m-derivation in the shunt sequence.

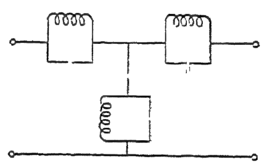


FIG. 191.—Structure which potentially realizes the index and characteristic impedance functions (960) and (960a) when the point of infinite attenuation lies *below* the critical frequency x_2 in z_0.

On the other hand, if $x_\infty > x_2$, the residue of z_b at $x = x_2$ is *less* than that of z_a. The ladder development again begins with a series branch consisting of an anti-resonant component which removes the pole at $x = x_2$ completely from z_b, leaving for z_b' a condenser and for z_a' the same form as z_a (an anti-resonant component). The functions y_a' and y_b', therefore, have a common pole at infinity, and since these functions cross at $x = x_\infty$, the residue in y_b' at infinity is *less* than that in y_a'. Hence y_b' may be removed completely, leaving $y_b'' = 0$ and y_a'' of the same form as y_a' and hence of the same form as z_a' or z_a. The resulting structure is that shown in Fig. 192, which is *not* a Zobel type.

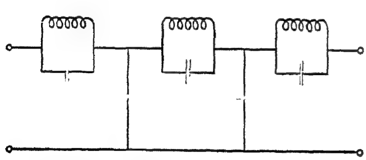


FIG. 192.—Structure which replaces that of Fig. 191 when the point of infinite attenuation lies *above* the critical frequency x_2 in z_0.

For more complicated z_0-functions it may be seen thus that the re-

sulting section is identifiable as a Zobel type whenever x_∞ lies *below* the critical frequencies of z_0. When $m > 1$, or when we have a pair of complex m-sections, the lattices are not reducible to the ladder form, although either series or shunt branches may be taken out as a first step in an attempted development, leaving a remainder in the form of a lattice structure. The theory of network equivalents unfortunately is not yet in a sufficiently satisfactory state of development to enable us to set down more specific methods of procedure. The reader is thrown upon his own ingenuity in attempting to find possible bridged-T or other more suitable structures.

PROBLEMS TO CHAPTER X

10–1. Consider the symmetrical low-pass constant-k filter, and show that if we convert to the equivalent lattice we find for the T-section

$$Z_a = \frac{jL_1\omega}{2};\ Z_b = \frac{jL_1}{2\omega}(\omega^2 - \omega_1^2)$$

$$y_0 = \sqrt{1 - \frac{\omega_1^2}{\omega^2}};\ z_0 = \sqrt{1 - \frac{\omega^2}{\omega_1^2}}$$

where ω_1 is used to denote the cut-off frequency; and for the Π-section

$$\frac{1}{Z_a} = \frac{jC_2}{2\omega}(\omega^2 - \omega_1^2);\ \frac{1}{Z_b} = \frac{jC_2\omega}{2}$$

$$y_0 = \sqrt{1 - \frac{\omega_1^2}{\omega^2}};\ z_0 = 1 \Big/ \sqrt{1 - \frac{\omega^2}{\omega_1^2}}.$$

Plot the functions y_0 and z_0 in each case vs. the ratio ω/ω_1 and show that by using the first relation (846) the resulting attenuation and phase functions agree with those found in the previous chapter. Draw the lattice structures and also sketch the lattice reactance functions Z_a and Z_b for each case.

10–2. With reference to the previous problem, show that the process of mid-series derivation may be expressed by multiplying the lattice reactance Z_a by m and dividing Z_b by m, and check the relations (753) of the previous chapter. Alternatively show that the shunt-derivation may be obtained by multiplying $1/Z_b$ by m and dividing $1/Z_a$ by m, and thus check the relations (772) or (772a). Plot the functions y_0 corresponding to the derived types for $0 < m < 1$ and also for $m > 1$, and sketch the corresponding attenuation and phase functions for each, assuming numerical values for m such as 0.6 and 1.2, for example, in order that the sketches may be quantitative with regard to their major characteristics. Also sketch the lattice reactances for $0 < m < 1$ and $m > 1$, and show that they are physical in both cases. Again make these sketches show quantitatively the essential features which give rise to the particular properties of the derived types.

10–3. Write the analytic expressions for the y_0-functions of a low-pass filter capable of meeting the following specifications:

(a) The attenuation function should have the form of γ_{1k}.
(b) " " " " " " " " γ_{1km}.
(c) " " " " " " " " $\gamma_{1k} + \gamma_{1km}$.
(d) " " " " " " " " $\gamma_{1k} + \gamma_{1km} + \gamma_{1kmml}$.

Write the analytic expressions for the z_0-functions of a low-pass filter capable of meeting the following specifications:

(a) The impedance function should have the form of W_{1k}.
(b) ” ” ” ” ” ” ” ” W_{1km}.
(c) ” ” ” ” ” ” ” ” $W_{1kmm'}$.
(d) ” ” ” ” ” ” ” ” $W_{2kmm'}$.

In each case sketch the function versus the ratio ω/ω_1,

(i) assuming that the parameters are determined on the Cauer basis,
(ii) ” ” ” ” ” ” ” ” Bode ” .

For the index functions sketch also the general form of the corresponding attenuation and phase functions.

10–4. For each of the index functions in Problem 10–3, combine with the z_0-function of part (a) and thus determine the analytic forms of the resulting lattice reactances. Repeat for combinations with the z_0-functions of parts (b), (c), and (d), and make sketches showing the form of the reactance functions (i) when the parameters in y_0 and z_0 are determined on the Cauer basis; (ii) when the parameters in y_0 and z_0 are determined on Bode basis: (iii) when the parameters in y_0 are determined on the Bode basis and those in z_0 on the Cauer basis.

10–5. For the y_0-functions of Problem 10–3 determine the parameters on the Cauer basis so that in

(b) the minimum attenuation is 30 db.
(c) ” ” ” ” 40 db.
(d) ” ” ” ” 50 db.

In each case determine the frequencies (in terms of ω_1) at which this minimum attenuation is first reached as well as the remaining frequencies at which minimum or infinite attenuation occurs.

10–6. For the z_0-functions of Problem 10–3 determine the parameters on the Cauer basis so that in (b) the tolerance does not exceed 5 per cent, and in (c) and (d) the tolerance does not exceed 2 per cent. In each case determine the coverage in percentage of the theoretical band width.

10–7. For the y_0-functions (b), (c), and (d) of Problem 10–3 determine the parameters on the Bode basis. Plot the corresponding functions. Determine and plot the attenuation and phase functions vs. ω/ω_1. For each phase function determine the average slope t_d as defined in section 9, and plot the ratio $y_0/j \tan \frac{\omega t_d}{2}$ over the transmission range (this is the same as the ratio Q'/Q). From these plots determine the approximate maximum percentage deviation from constancy in the slope of the phase characteristic for the function

(b) over 70 per cent of the theoretical transmission band.
(c) ” 80 ” ” ” ” ” ” ” .
(d) ” 90 ” ” ” ” ” ” ” .

10–8. Combining the y_0-functions (c) and (d) of Problem 10–5 with the z_0-function (a) of Problem 10–3, synthesize on the basis of the points of infinite attenuation so as to obtain a cascade of simpler lattices. Show that when each component lattice corresponds to a Zobel derived type with $0 < m < 1$ it may be converted into an

equivalent T-section by the process of ladder development. Show that the Cauer parameter determinations for the y_0-functions fulfill this condition, and obtain the complete ladder development for each of the filters defined in this problem. Show that this would not be possible if the parameters in the y_0-functions were determined on the Bode basis.

10–9. Combining the y_0-functions (*c*) and (*d*) of Problem 10–7 with the z_0-function (*a*)·of Problem 10–3, synthesize on the basis of the points of infinite attenuation so as to obtain a cascade of simpler lattices. For each of these component lattices carry a ladder development as far as possible and thus find a structure for each of the two filters defined in this problem. Alternatively attempt a conversion of each component lattice into an equivalent bridged-T network. As a third alternative, attempt a ladder development for the lattice reactances representing each complete filter, and also attempt a conversion of the remaining undeveloped portion into the equivalent bridged-T form.

10–10. Combine the y_0-function (*c*) of Problem 10–5 with the z_0-function (*b*) of Problem 10–6, and determine the resulting lattice reactance functions. Carry through a ladder development realizing the filter thus specified, and find all possible alternative equivalent developments.

10–11. A low-pass filter designed on the lattice basis involves the following index and characteristic impedance functions respectively

$$y_0 = \frac{\sqrt{\omega_1^2 - \omega^2}}{j\omega m}; \quad z_0 = \frac{H\sqrt{\omega_1^2 - \omega^2}}{(\omega_2^2 - \omega^2)}.$$

The parameter m is real and lies between zero and unity so that a point of infinite attenuation occurs at a real frequency which we shall denote by ω_∞. For the two cases·

(*a*) $\omega_2 < \omega_\infty < \infty$
(*b*) $\omega_1 < \omega_\infty < \omega_2$

(i) Sketch the lattice reactances z_a and z_b as functions of frequency showing the relative locations of the critical frequencies with respect to the point at which infinite attenuation occurs. What are the relative magnitudes of the residues in the pole at $\omega = \omega_2$ for the two reactances, and how does this affect the plots of these functions?

(ii) Starting with partial fraction expansions of z_a and z_b, carry through the process of ladder development with sufficient care so that the form of the resulting network in each case may be recognized.

(iii) Point out the total number of network elements necessary for the realization of each case.

(iv) Which of the two cases falls into the category of a Zobel type? How would you designate the resulting structure in the latter terminology?

10–12. If a low-pass filter is specified to have an index function of the form given in the preceding problem with its parameter determined on the Cauer basis, but an impedance function of the general form

$$z_0 = \frac{H\sqrt{\omega_1^2 - \omega^2}\,(\omega_3^2 - \omega^2)\cdots}{(\omega_2^2 - \omega^2)(\omega_4^2 - \omega^2)\cdots},$$

show that, if the single frequency of infinite attenuation is less than ω_2, the residues in the poles of z_a are *all* smaller than those in the coincident poles of z_b, and hence that the ladder development may be completed after one step. Show that then the resulting structure may be identified with a Zobel derived type. Show that if the frequency of infinite attenuation is larger than ω_2, the filter is no longer of the

Zobel type and that further steps are necessary in completing the ladder development. Also show that the Cauer-designed lattice is identifiable as a Zobel type only when its y_0-function contains a *single* parameter.

10–13. A low-pass filter is designed with the y_0-function

$$y_0 = \frac{j\omega H\sqrt{\omega_1^2 - \omega^2}\,(\omega_b^2 - \omega^2)}{(\omega_a^2 - \omega^2)}.$$

Sketch the general character of the attenuation and phase functions in the attenuation and transmission ranges respectively for

(*a*) parameters determined by the Bode method,
(*b*) " " " " Cauer "

What is the total phase shift throughout the transmission range in either case? Sketch curves of delay-time versus frequency for these cases, and also discuss relative effects of incidental dissipation, assuming the latter to be uniform.

10–14. In the design of a low-pass filter on the lattice basis, the following index and characteristic impedance functions are chosen

$$y_0 = \frac{h\omega\sqrt{\omega^2 - \omega_1^2}}{(\omega^2 - \omega_a^2)}; \quad z_0 = \frac{\sqrt{\omega_1^2 - \omega^2}}{\omega_1}.$$

The cut-off frequency is ω_1, while $\omega_a < \omega_1$ and h are design parameters for the index function.

(*a*) Following the rule of the half-spaced cut-off factor on the Bode basis, find ω_a in terms of ω_1. Determine the average slope t_d of the phase characteristic, and find that value of h for which the tangent to the phase characteristic at the origin coincides with this average slope. Will the filter thus determined have a peak of infinite attenuation at a real frequency in the attenuation range? Sketch the resulting lattice reactances z_a and z_b.

(*b*) Using the same index and characteristic impedance functions, determine the parameters in y_0 so that this function approximates unity in the attenuation range in the Tschebyscheff manner with a maximum tolerance of 2 per cent, thus giving rise to a minimum attenuation of 40 db. How many peaks of infinite attenuation will the filter then have? Sketch the resulting lattice reactances in this case.

10–15. A symmetrical filter is specified in terms of the reactances of the equivalent lattice structure; they are

$$z_a = \frac{\lambda^2 + 0.64}{\lambda},$$

$$z_b = 0.96\lambda\,\frac{(\lambda^2 + 1)}{\lambda^2 + 0.64}.$$

Consider the possibilities of building this network (*a*) as a structure containing mutual inductances but no ideal transformers, (*b*) as a ladder structure with no mutual inductances. (Give the arrangement of the elements, but do not calculate their exact magnitudes.)

10–16. A reactive network having the following open- and short-circuit impedances measured from one end is desired

$$Z_{oc} = 4.5\lambda + \frac{\lambda}{\lambda^2 + 1} + \frac{2}{\lambda},$$

$$Z_{sc} = 4.5\lambda + \frac{9\lambda}{3\lambda^2 + 1}.$$

(*a*) Show that such a network is physically realizable;

(*b*) Find a structure (preferably simple) realizing these characteristics.

10–17. Design the filter called for in Problem 9–12 on the lattice basis, and compare the resulting structure with that obtained by the conventional methods.

10–18. Repeat Problem 10–17 for the filter called for in Problem 9–13.

10–19. Design on the lattice basis the filter called for in Problem 9–16. Note that this requires a modification of the Tschebyscheff parameter determination on a cut-and-try basis.

10–20. Determine network realizations according to Figs. 179 and 181 for the filters specified in Problem 10–8.

10–21. Repeat Problem 10–20 for the filters specified in Problem 10–9.

10–22. Repeat Problem 10–20 for the filter specified in Problem 10–10.

10–23. Design a low-pass filter with theoretical cut-off at 3500 cycles per second. The characteristic impedance may be of the constant-k type. The slope of the phase function should not deviate from a constant value by more than 10 per cent over 90 per cent of the theoretical transmission range, and the attenuation function should reach 40 db at 4000 cycles per second and should not fall below 50 db for frequencies above 5000. Determine the lattice reactances, and synthesize on the basis of the points of infinite attenuation. Alternatively attempt a ladder development and conversion of the undeveloped remainder.

10–24. In the filters of Problems 10–8, 10–9, and 10–23 assume R/L ratios equal to 100 and negligible condenser losses. Thus determine approximately the attenuation function throughout the transmission ranges due to the incidental dissipation.

10–25. If in the filters of Problems 10–8 and 10–9 it is specified that the incidental attenuation in the transmission range shall not vary by more than 3 db, what must be the uniform R/L ratio for the coils (assuming zero condenser loss) in the Cauer and Bode designs?

10–26. For the Cauer filters of Problem 10–8 assume terminal resistances equal to the nominal value; determine and plot the insertion loss for $1 < \omega/\omega_1 < 4$, and compare with the attenuation loss.

10–27. Repeat Problem 10–26 for the Bode filter designs specified in Problem 10–9.

10–28. By means of network modifications involving the ideal transformer as illustrated in Fig. 187, convert the symmetrical filter designs of Problem 10–8 into dissymmetrical ones meeting a ratio of input-to-output impedance levels of 1 : 5.

10–29. By means of the transformations indicated in eqs. (916) to (919*a*), convert the low-pass designs specified in Problems 10–8 and 10–9 into the band-pass class.

10–30. Design a pair of low-pass lattices with constant-k characteristic impedances and index functions given by y_k/m and $y_k/\bar{m}$ respectively, m and $\bar{m}$ being conjugate values. Let the m's be determined so that a point of infinite attenuation for the dissipative structure falls at the cut-off frequency ω_1. Let the uniform R/L ratio be equal to 100. Plot the individual attenuation functions and their sum. Let the dissipation become zero, and note the effect on the resulting attenuation functions. What form has the phase function of this filter? Make plots extend over the range $0 < \omega/\omega_1 < 2.5$.

CHAPTER XI

THE TRANSIENT BEHAVIOR OF FILTERS

1. **The Fourier integral.** When a filter transmits a given time function to the exclusion of others, it does so on the basis of frequency discrimination. The basic idea involved in this process is so intimately associated with that leading to the Fourier integral representation of an arbitrary function that we shall find the discussion of the latter problem a very instructive as well as useful means for introducing our present topic.

A method of converting a time function into an equivalent frequency function, with which we are already familiar, involves the idea of expansion in a Fourier series. In order for such a conversion to lead to a representation valid for all time from minus to plus infinity, it is necessary that the time function be periodic over this same infinite time interval. Such a function has neither beginning nor end, and, therefore, is never met with-in practice. The reason why we nevertheless find good use for Fourier series representations lies in the fact that the requirement regarding the infinite time interval may usually be greatly modified without causing serious error in the results obtained.

As we know from our study of lumped-constant network behavior these results apply only to that portion of the response designated as the steady state. Fourier series analysis thus is useful only in connection with problems where a knowledge of the steady-state behavior alone is sufficient. During a certain interval of time after the application or removal of a periodic force, the response of the network is influenced by its natural or force-free behavior. Owing to the dissipative character of the network, these transient intervals, although theoretically of infinite duration, are for all practical purposes not only finite but usually of relatively short duration. However, when we are particularly interested in the manner in which the steady state builds up or down, then the steady-state analysis, as it has so far been interpreted, gives us no useful information.

The reason for this lies in the fact that the Fourier series does not accurately describe the actual force function because it contains no information regarding its beginning and end; i.e., it does not account for its transient nature. On the contrary, it carries with it the tacit assumption that this force has been applied since $t = -\infty$ and will con-

tinue to be applied until $t = +\infty$. The Fourier series solution, therefore, can give no account of the transient intervals.

If it were possible to remove this restriction, the Fourier method of analysis for linear systems would obviously become far more useful. In that case the response of a network as a result of the application of a transient force function (one with a beginning and end) could be determined by the simple process of superposing steady-state solutions whose formulation involves much less labor, as we are well aware. It is the object of our present discussion to show that this can be done, and that the result of this process, although applicable to any linear passive network, is particularly useful in enabling us to arrive at certain characteristic properties regarding the response of filter networks. In addition to this, the present analysis gives us a new and extremely useful point of view regarding network response which is free from the distinction between transient and steady-state behavior and consequently free from the restrictions which are part of such an interpretation.

The reader who is unacquainted with this subject will doubtless get a clearer picture of the fundamental ideas involved from a heuristic discussion than from a mathematically rigorous one. Therefore, we shall place clarity before rigor, recognizing that rigor can be appreciated only after sufficient familiarity with the basic ideas has been gained. For the rigorous demonstration of some of the more mathematically involved questions, the reader may subsequently refer to numerous other sources.

Since the Fourier integral is an outgrowth of the Fourier series, we begin with the latter. This we shall again consider in the complex form which has proved so far superior to the sine and cosine form in its application to network analysis. We recall that if $f(t)$ be a periodic function[1] defined for $-\infty < t < \infty$, with a finite period $\tau = 2\pi/p$, then its Fourier series representation may be written as

$$f(t) = \sum_{\nu=-\infty}^{\infty} a_\nu e^{j\nu p t}, \tag{961}$$

where[2]

$$a_\nu = \frac{p}{2\pi}\int_{-\frac{\pi}{p}}^{\frac{\pi}{p}} f(t)\cdot e^{-j\nu p t}dt \tag{961a}$$

are usually designated as the Fourier coefficients, but for our present purpose are more appropriately interpreted as functions of the discon-

[1] With certain restrictions which need not be considered at the moment.

[2] The limits on this integral must extend over a period but are otherwise arbitrary.

tinuous variable νp, where ν takes on all positive and negative integer values inclusive of zero. The series (961) is thus a conversion from the frequency function a_ν, or $a(\nu p)$, to the time function $f(t)$, while the integral (961a) is the inverse. Although t is a continuous variable and νp a discontinuous one, this pair of relations already makes evident a certain inherent symmetry. Usually $f(t)$ is a real function and a_ν complex, in which case a_ν and $a_{-\nu}$ are conjugates. These may be written as $|a_\nu|e^{j\psi_\nu}$. The functions $|a_\nu|$ and ψ_ν may be represented graphically as functions of νp by means of ordinates erected at points corresponding to integer ν's extending from $-\infty$ to ∞. These plots are spoken of as **line spectra,** more specifically as the **amplitude** and **phase spectra** corresponding to $f(t)$. Thus the time function $f(t)$ is replaced by one complex frequency spectrum or by a pair of real frequency spectra.

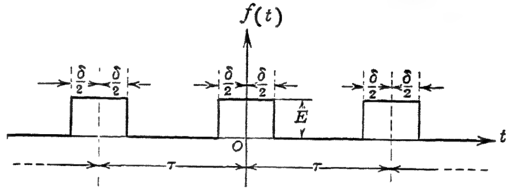


FIG. 193.—Periodic function consisting of a succession of rectangular uni-directional impulses.

Although these spectra are discontinuous, they are infinite in extent, the frequency increment between lines being uniform and equal to p. As an illustration let us consider the function shown in Fig. 193. This consists of similar square impulses of duration δ seconds and amplitude E at intervals of τ seconds, with the time-origin chosen at the center of one such impulse. Applying (961a) we have

$$a_\nu = \frac{Ep}{2\pi}\int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} e^{-j\nu pt}dt = \frac{\delta Ep}{2\pi}\cdot\left(\frac{\sin \nu p\frac{\delta}{2}}{\nu p\frac{\delta}{2}}\right). \tag{962}$$

These coefficients are real, so that substitution into (961) leads to a cosine series. This is to be expected since the function $f(t)$ is even.

The factor $\delta Ep/2\pi$ is constant. The variation of a_ν is thus given by the second factor in (962) which is of the form

$$\frac{\sin u}{u}.$$

A plot of (962), considering νp as a continuous variable, thus forms the envelope which limits the lines for the spectrum representation of a_ν. These lines are erected at intervals of p to the right and left, inclusive of the origin. Fig. 194 shows this line spectrum for the specific case $\tau = 5\delta$ or $p = 2\pi/5\delta$. The envelope is shown dotted. This has zeros

at intervals of $\nu p = 2\pi/\delta$ which are independent of the ratio of δ to τ. The spacing of the lines, however, definitely depends upon this ratio. In the present example the fifth and multiples of the fifth harmonic fall at the zeros of the envelope function and hence are zero. The amplitudes of a_ν are allowed to have both positive and negative values. This procedure can be varied by considering all amplitudes positive and supplying the negative components with phase angles of π radians. In our example the accompanying phase spectrum is everywhere zero and hence need not be considered.

The envelope function shows the variation in relative magnitude of the harmonic components. We note that the components in the vicinity of the origin are most important, the largest being the zero frequency or constant component. Inspection of Fig. 194, therefore, makes it plausible that a facility which transmits faithfully all frequencies (in-

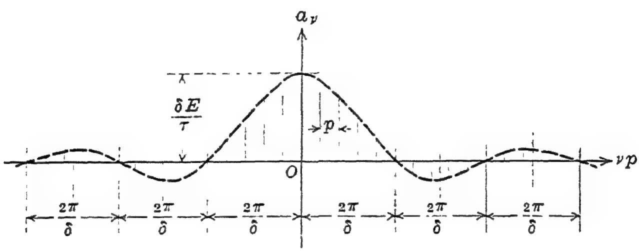


FIG. 194.—Line spectrum corresponding to the time function of Fig. 193 for the condition $\tau = 5\,\delta$.

cluding zero) up to, say, $\nu p = 2\pi/\delta$ or $3\pi/\delta$ ought to do fairly well in the way of reproducing the time-function of Fig. 193. Thus by converting a given time-function into an equivalent frequency function (conversion of Fig. 193 into Fig. 194) we are able to estimate by inspection whether a given facility, as a transmitting medium for the time function, is good, fair, or poor, on the basis of its frequency response characteristics.

So far the process is limited to periodic time functions of infinite extent, however. In order to come closer to our ultimate goal, let us suppose that, instead of an infinite periodic repetition of impulses, we have to transmit only a single one. Considering the plot of Fig. 193, the thought immediately comes to us that such a time function results from our present $f(t)$ merely by allowing the period τ to become indefinitely large. In the limit we shall be left with the symmetrically located impulse (of finite duration δ) at the origin. Heuristically we see no reason why the Fourier series should not represent the case for any τ however large, and finally in the limit $\tau \to \infty$ also. Let us see

what effect such an increase in τ has upon the frequency spectrum for our function. If the spectrum still exists in the limit then the principle of the Fourier series also holds good.

Noting that the envelope function (considering νp as the independent variable) does not depend upon τ except that its amplitude is inversely proportional to it, we see that if we double τ (for example) and take the unit lengths for ordinates twice as large, the dotted curve of Fig. 194 remains unchanged. The variation of harmonic amplitudes with frequency, therefore, remains the same. The spacing of the lines, however, becomes half as large, since p varies inversely as τ. The fundamental frequency also becomes half as large; i.e., the first line is nearer the origin by a factor $\frac{1}{2}$. If we proceed to double and redouble τ, modifying the scale for ordinates each time, the appearance of the spectrum changes only in that the lines move more closely together. In the limit $\tau \to \infty$ it will take an infinitely large scale-factor in order to keep the plotted amplitudes finite, and the spacing between lines will be zero. Another way of putting this is to say that all amplitudes become differential in magnitude (but retain the same frequency variation!), and the spacing between lines also becomes differential. On account of the differential spacing, all frequencies are present. The spectrum is said to be *continuous*.

This transition may be formulated mathematically in the following way. Substituting (962) into (961) we have

$$f(t) = \frac{\delta E}{2\pi} \sum_{\nu=-\infty}^{\infty} \left(\frac{\sin \nu p \frac{\delta}{2}}{\nu p \frac{\delta}{2}} \right) \cdot \Delta(\nu p) \cdot e^{j\nu p t}, \tag{963}$$

where $\Delta(\nu p)$ is written in place of p for the uniform frequency increment. As τ is allowed to increase, p decreases, so that we must let ν increase in such a way that the quantity νp may refer to any frequency within the spectrum as before. This is the same as to say that, for double the value τ, a given frequency component doubles its order. In the limit $\tau \to \infty$, νp becomes a continuous variable which we shall denote by ω. The finite frequency increment $\Delta(\nu p)$ then becomes the differential increment $d\omega$. In symbols we have

$$\left.\begin{aligned} \tau &\to \infty \\ p &\to 0 \\ \nu &\to \infty \\ \nu p &\to \omega \\ \Delta(\nu p) &\to d\omega \end{aligned}\right\} \tag{964}$$

When the increments are of differential magnitude, the sum becomes an integral, so that in the limit we have for our new $f(t)$

$$f(t) = \frac{\delta E}{2\pi} \int_{-\infty}^{\infty} \left(\frac{\sin \omega \frac{\delta}{2}}{\omega \frac{\delta}{2}} \right) \cdot d\omega \cdot e^{j\omega t}. \tag{965}$$

It can be demonstrated rigorously that this function actually represents the single square-topped impulse in the Fourier sense. This is not done by justifying the various steps which we have taken here in our heuristic derivation, but rather by assuming the result (965) and showing that it does represent the function as we have described it. This demonstration adds nothing to the general picture which we have given, however, and for our purposes may be omitted.

The result (965) is spoken of as the **Fourier integral** corresponding to the single, symmetrically located, flat-topped impulse of duration δ and magnitude E. The value of the integrand for a particular frequency is still referred to as a frequency component. In fact, the significance, physical interpretation, and use as applied to network response remain exactly the same as for the series. The only differences are that the components have differential magnitudes, are differentially spaced, and that we deal with an integral instead of a sum.

The result (965) in more general terms is written as

$$f(t) = \int_{-\infty}^{\infty} g(\omega) \cdot d\omega \cdot e^{j\omega t}, \tag{965a}$$

where for our particular example (the single Morse dot) we have

$$g(\omega) = \frac{\delta E}{2\pi} \left(\frac{\sin \omega \frac{\delta}{2}}{\omega \frac{\delta}{2}} \right). \tag{965b}$$

Thus the present result is exactly analogous to the series (961) where instead of the coefficients a_ν we have the continuous function $g(\omega) \cdot d\omega$ which is clearly of differential magnitude. In discussing the variation of amplitudes with frequency we consider the function $g(\omega)$ which has finite amplitudes and in our example has the same general shape as the dotted envelope of Fig. 194. The situation for the single Morse dot is summarized in Fig. 195 in which part (a) shows the time function and part (b) the corresponding frequency function. Both functions are defined for the interval $-\infty$ to ∞ for t and ω respectively.

This result is graphically interpreted to the effect that if all the harmonic components with relative amplitudes indicated by $g(\omega)$ were drawn as functions of time, and ordinates added, the single flat-topped impulse would result. Since each harmonic component extends from $t = -\infty$ to $t = \infty$, this means that for $t < -\frac{\delta}{2}$ and $t > \frac{\delta}{2}$ the linear superposition of harmonic components results in a complete cancelation, while for the interval $-\frac{\delta}{2} < t < \frac{\delta}{2}$ they add to give the constant value E. An appreciation of how this comes about is gained by referring to the discussion of a finite frequency group in section 6 of Chapter III.

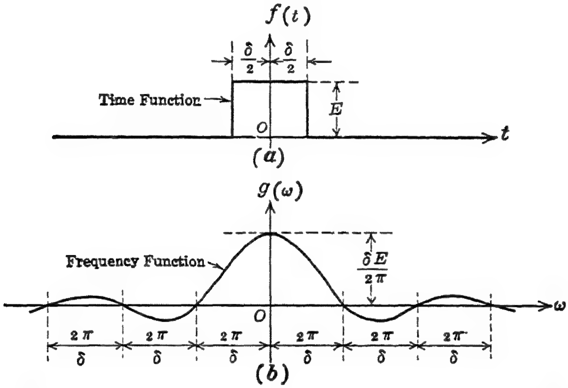


FIG. 195.—Time function in the form of a single Morse dot (*a*), and its frequency transform (*b*). This pair of functions represents those of Figs. 193 and 194 for the limit $\tau \to \infty$.

In the present instance the mean frequency is zero, so that the result for the frequency group of Fig. 21, page 100, is given by the envelope function (186) alone as illustrated in Fig. 22. As pointed out there, the distance[1] between the hump at the origin and the next nearest humps to the right and left is inversely proportional to the spacing between lines in the group spectrum. When the latter becomes continuous, the resultant consists of only a single hump at the origin. This is then given by (186*a*), which, as we pointed out, is the limiting form for $n \to \infty$. Since the amplitudes in the group of Fig. 21 are all taken equal, the resultant is not a rectangular impulse like Fig. 195(*a*) but instead has exactly the shape of Fig. 195(*b*). Had an amplitude distribution like part (*b*) been assumed for the frequency group, the result would have been rectangular like part (*a*). We have here a special case of a very

[1] In the discussion in section 6 of Chapter III the independent variable is distance, but the idea applies to any other variable, for example, time, as well.

interesting general property of Fourier integral transformations, namely, that the variables t and ω are interchangeable in the sense that: Whereas the time function of part (*a*) of Fig. 195 has as its mate the frequency function of part (*b*), a time function like part (*b*) has as its mate a frequency function like part (*a*). This we shall formulate more generally in the following.

Although it is possible to derive the Fourier integral for almost any function by considering the periodic extension of that function first and then applying the limiting process to the resulting series as was done in the above example, this is quite unnecessary if we formulate the time integral and its frequency mate[1] in general terms similar to the pair of relations (961) and (961*a*) for the series. Thus, if in these expressions we consider the limiting process (964), the following forms are obtained for the general case of any non-periodic (transient) time function with which we shall have to deal in practice

$$\left.\begin{aligned} f(t) &= \int_{-\infty}^{\infty} g(\omega)d\omega e^{j\omega t} \\ g(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)dte^{-j\omega t} \end{aligned}\right\} \qquad (966)$$

The restrictions on $f(t)$ may be stated by saying that:

(*a*) *The function* f(t) *must be continuous over intervals of finite length. At points of discontinuity, say* t = a, *the value of the function shall be understood to be the arithmetic mean, i.e.,*

$$f(a) = \tfrac{1}{2}\{f(a-0) + f(a+0)\}. \qquad (967)$$

(*b*) *The integral of the magnitude of* f(t) *over the region* $-\infty < t < \infty$ *must be finite, i.e.,*

$$\int_{-\infty}^{\infty} |f(t)|dt \qquad \textit{must exist.} \quad (967a)$$

Any practical function fulfils these conditions.

The pair of relations (966) are more symmetrical than (961) and (961*a*) for the periodic function. This symmetry may be enhanced by writing the factor $1/\sqrt{2\pi}$ before each integral, which may evidently be done instead of having no factor for the integral of $f(t)$ and the factor $1/2\pi$ for that of $g(\omega)$. Although many writers do this, we shall not bother with it here since it merely amounts to a change in the scale factor,

[1] For a very complete table of functions and their Fourier transforms the reader is referred to "Fourier Integrals for Practical Applications," by G. A. Campbell and R. M. Foster, Bell Telephone System Technical Publication, Monograph B-584.

and in our opinion the pair of relations (966) are sufficiently symmetrical without this change.

When $f(t)$ is *real* we may draw the following general conclusions from (966)

(a) *If* $f(t)$ *is an even function of* t, *then* $g(\omega)$ *is real and an even function of* ω.

(b) *If* $f(t)$ *is an odd function of* t, *then* $g(\omega)$ *is a pure imaginary and an odd function of* ω.

(c) *When* $f(t)$ *is neither even nor odd it can be written as* $f(t) = f_1(t) + f_2(t)$ *where* $f_1(t)$ *is even and* $f_2(t)$ *odd. Then* $g(\omega)$ *is complex; i.e.,* $g(\omega) = g_1(\omega) + jg_2(\omega)$ *and*

$$g_1(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_1(t)dte^{-j\omega t};\ g_2(\omega) = \frac{1}{2\pi j}\int_{-\infty}^{\infty} f_2(t)dte^{-j\omega t}$$

Here $g_1(\omega)$ *is even and* $g_2(\omega)$ *odd as required by the obvious fact that* $g(-\omega)$ *is the conjugate of* $g(\omega)$.

According to the conclusion (a), therefore, we see that if $f(t)$ is an even function, the sign of ω in the exponent of e in the second integral (966) may just as well be changed from minus to plus. In this case the integrals are completely similar, which means that t and ω may be interchanged; i.e., the amplitude spectrum if considered as a time function has the original time function as its amplitude spectrum. Parts (a) and (b) of Fig. 195 bear this relation to one another, as already pointed out.

In the general case (c) where $g(\omega)$ is complex, we may write it as

$$g(\omega) = G(\omega)e^{j\psi(\omega)}, \tag{968}$$

in which $G(\omega)$ represents the magnitude of $g(\omega)$. In this form $G(\omega)$ is called the *amplitude spectrum* and $\psi(\omega)$ the *phase spectrum* corresponding to $f(t)$. Cases (a) and (b) are then defined by $\psi(\omega) = 0$ and $\psi(\omega) = \pi/2$ respectively. In general, $G(\omega)$ is an *even* and $\psi(\omega)$ an *odd* function.

The relations (966) are known in mathematics as **Fourier transforms.** They have a number of useful properties of which the following will be of particular interest to us. If in the first integral we

$$\text{replace } g(\omega) \text{ by } g(\omega)\cdot e^{\pm j\omega t_0} \tag{969}$$

then

$$f(t) \text{ is replaced by } f(t \pm t_0). \tag{969a}$$

This means that the addition of a linear phase function (with slope $= t_0$) to $\psi(\omega)$ amounts to a displacement of the origin in the corresponding time function, and vice versa.

If in the second integral we

$$\text{replace } f(t) \text{ by } f(t)\cdot e^{\pm j\omega_0 t} \tag{970}$$

then

$$g(\omega) \text{ is replaced by } g(\omega \pm \omega_0). \tag{970a}$$

In order to interpret the change expressed by (970) we note that if $f(t)$ is a real function then $\mathcal{R}e\left[f(t)\cdot e^{j\omega_0 t}\right]$ represents a cosine function of frequency ω_0 whose amplitude is modulated by $f(t)$. Since the response of a network to the force $f(t)\cdot e^{-j\omega_0 t}$ is the conjugate of that to $f(t)\cdot e^{j\omega_0 t}$, we may consider the latter as the driving force and take only the real part of the result, just as we have done in our previous network analysis. Thus the change expressed by (970) amounts to superposing the time function $f(t)$ (as an envelope) upon a carrier frequency ω_0. The result (970a) then states that this merely has the effect of displacing the frequency spectrum by an amount ω_0. This rather interesting result places the function of modulation (and demodulation) in a very clear light and shows its generality with respect to any form of modulation envelope.

2. Relation to contour integrals. In network analysis we frequently wish to consider a force function which is zero before $t = 0$ (designated as the switching instant) and constant thereafter. Such a function, of course, is not physically realizable, but useful for investigating network behavior. For this case the condition (967a) is evidently not fulfilled so that the Fourier integral representation cannot be carried through. In order to overcome this difficulty the function may be modified in such a way that it satisfies the condition (967a) and yet is near enough to the desired form to retain its usefulness in the problem to be treated. This may be accomplished in several ways. A modification which not only meets the situation but at the same time clarifies the relation between the complex Fourier integral and the corresponding contour integral (as defined in function theory) proceeds by defining the function as

$$\left.\begin{aligned} f(t) &= 0 \quad \text{for} \quad t < 0 \\ f(t) &= Ee^{-\alpha t} \quad \text{for} \quad t > 0 \end{aligned}\right\}, \tag{971}$$

in which α is a real positive number. By considering α sufficiently small, the usefulness of the function is unimpaired while the condition (967a) is satisfied and hence the Fourier integral representation made possible. We shall see that this is true even though we allow α to approach zero provided this limit is properly interpreted with regard to the evaluation of the Fourier integral.

Applying (966) to the time function defined by (971) we get

$$g(\omega) = \frac{E}{2\pi}\int_0^\infty e^{-(\alpha + j\omega)t}\,dt = \frac{E}{2\pi(\alpha + j\omega)}. \tag{972}$$

The result is finite because the integral vanishes at infinity. Substituting into the first relation (966), the corresponding time function becomes

$$f(t) = \frac{E}{2\pi}\int_{-\infty}^{\infty} \frac{e^{j\omega t}}{\alpha + j\omega}\,d\omega. \tag{973}$$

The desired time function is the limit of this integral as α approaches zero. If we simply put $\alpha = 0$, the integral becomes improper, since the

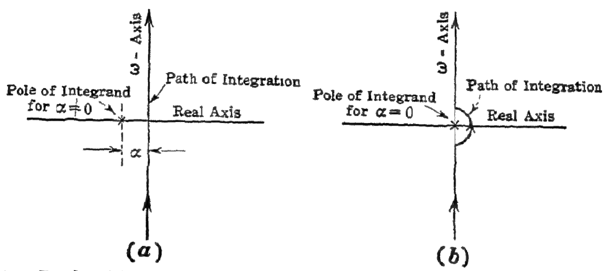


FIG. 196.—Path of integration (a) for the function defined by the integral (973) and (b), the manner in which this path must be modified in the limit $\alpha \to 0$.

integrand then has a pole at the origin which is on the path of integration, the latter being the axis of imaginaries. So long as α has a positive value, however small, the pole lies on the real axis, α units to the left of the origin, so that integration along the imaginary axis avoids this pole, thus keeping the integral proper. This situation tells us that if we wish to make α zero we must in the limit carry our path of integration so as to avoid the origin by passing slightly to the *right* of it, i.e., by distorting the path in the vicinity of the origin in the manner illustrated in Fig. 196 (a) and (b) by means of a small semicircular bulge the radius of which may subsequently be allowed to approach zero.

So long as we keep in mind the manner in which the path of integration in the vicinity of the origin is to be treated, we may put $\alpha = 0$ in (973) without causing trouble. The truth of this may be demonstrated in either of two ways. With $\alpha = 0$, (973) may be written

$$f(t) = \frac{E}{2\pi j}\int_{-\infty}^{\infty} \frac{\cos \omega t}{\omega}\,d\omega + \frac{E}{2\pi}\int_{-\infty}^{\infty} \frac{\sin \omega t}{\omega}\,d\omega. \tag{973a}$$

The second of these integrals is proper and needs no special comment. The first is the one whose integrand has a pole at the origin. Let us

split the path of integration for this integral into three parts as indicated by

$$\int_{-\infty}^{\infty} = \int_{-\infty}^{-\rho} + \int_{-\rho}^{\rho} + \int_{\rho}^{\infty}, \tag{974}$$

where ρ is a small quantity which we shall consider equal to the radius of the semicircle in Fig. 196 (b). Since the integrand in the first integral of (973a) is an odd functiốn of ω and vanishes at infinity, it is clear that the first and last integrals (974) cancel each other so that we have

$$\frac{E}{2\pi j}\int_{-\infty}^{\infty} \frac{\cos \omega t}{\omega}\, d\omega = \frac{E}{2\pi j}\int_{-\rho}^{\rho} \frac{\cos \omega t}{\omega}\, d\omega. \tag{975}$$

Having specified that ρ is a vanishingly small quantity, it is proper, for the integration (975), to replace $\cos \omega t$ by unity and get for any finite t

$$\frac{E}{2\pi j}\int_{-\rho}^{\rho} \frac{\cos \omega t}{\omega}\, d\omega \rightarrow \frac{E}{2\pi j}\int_{-\rho}^{\rho} \frac{d\omega}{\omega}. \tag{975a}$$

According to our plan to avoid the origin along a semicircle of radius ρ we now put $j\omega = \rho e^{j\varphi}$, whence $jd\omega = j\rho e^{j\varphi} d\varphi$, and since we are to pass the origin on the right[1] the limits become $-\pi/2$ to $\pi/2$, so that the integral becomes

$$\frac{E}{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\varphi = \frac{E}{2}. \tag{975b}$$

This is independent of ρ, and so letting $\rho \rightarrow 0$ does not affect the result. We thus have for our time function (973a)

$$f(t) = \frac{E}{2} + \frac{E}{2\pi}\int_{-\infty}^{\infty} \frac{\sin \omega t}{\omega}\, d\omega. \tag{973b}$$

The infinite integral

$$\int_{-\infty}^{\infty} \frac{\sin \omega t}{\omega}\, d\omega$$

is equal to the finite area under the curve

$$\frac{\sin \omega t}{\omega}; \quad -\infty < \omega < \infty,$$

which has the familiar shape of Fig. 195(b) for $t > 0$ and the negative of this for $t < 0$. The magnitude of this area is found equal to π, so

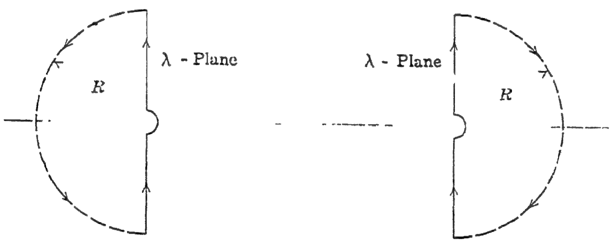
[1] This follows from the fact that, if we pass on the right, $d\varphi > 0$, i.e., φ *increases*. If we pass on the left, φ decreases, and the limits would be $3\pi/2$ to $\pi/2$. This changes the algebraic sign of the result.

that (973b) is seen to represent the function as defined by (971) for $\alpha = 0$. At $t = 0$ the integral behaves in the Fourier sense; i.e., it gives rise to the result $f(0) = E/2$.

Another method of treating this situation makes use more completely of the principle of integration in the complex plane. Since the Fourier integral and its mate hold generally for complex arguments, we may put $\lambda = j\omega$ in (973) and consider λ as a *complex variable.* In a sense this was already done in the above procedure for the vicinity of the origin. Here we shall utilize more fully the properties of functions of a complex variable with regard to their integration along a closed contour. Setting $\alpha = 0$, the so-called **contour integral** equivalent to (973) becomes

$$f(t) = \frac{E}{2\pi j}\int_{-j\infty}^{j\infty} \frac{e^{\lambda t}}{\lambda}\,d\lambda, \tag{976}$$

the integration again being along the axis of imaginaries so as to avoid the pole at the origin as illustrated in Fig. 196(b).



Path for t Positive
(a)

Path for t Negative
(b)

FIG. 197.—Paths of integration for $t>0$, (a), and $t<0$, (b), for the contour integral (976) corresponding to the Fourier integral (973).

Since this is not a closed contour, the principles of contour integration are not directly applicable. We note, however, that the integrand is regular except at the origin where it has a simple pole. This suggests that the path may be artificially closed by means of a large semicircle. In order that the integrand be negligible on this semicircle, it must be so drawn that λt has a *negative* real part. Thus the existence of the integral requires that for $t < 0$ the semicircular return path be drawn in the right-half λ-plane, and for $t > 0$ in the left-half λ-plane. The resulting closed contours for positive and negative t are shown in Fig. 197(a) and (b), respectively, R being the radius of the semicircle. The path for $t < 0$ encloses no singularities, and hence the value of the integral is zero. For $t > 0$ the path encloses the pole at the origin whose residue is unity so that we get $f(t) = E$ as required.

In order to prove that the return paths contribute nothing, we let

$\lambda = Re^{j\varphi}$. Then the integral on the return path for $t > 0$ (for example) becomes

$$j\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} e^{Rt(\cos\varphi + j\sin\varphi)}\, d\varphi,$$

or

$$j\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \{\cos(Rt\sin\varphi) + j\sin(Rt\sin\varphi)\} e^{Rt\cos\varphi}\, d\varphi.$$

Here

$$e^{Rt\cos\varphi} \cdot \sin(Rt\sin\varphi)$$

is an odd function of φ about $\varphi = \pi$ as well as about $\varphi = 0$. Its integral between symmetrical limits vanishes. This leaves

$$j\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} e^{Rt\cos\varphi}\cos(Rt\sin\varphi)\, d\varphi.$$

Since $Rt\cos\varphi$ is negative, and the path of integration finite (namely $= \pi$), R can be made large enough so that the value of this integral becomes negligible.

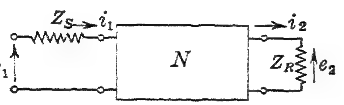


FIG. 198.—Schematic representation of a linear passive network with its associated terminal impedances and its excitation (e_1) and response (e_2, i_2) functions.

There are other ways of treating this case, such as finding the Fourier integral for the function which equals E only for the interval $0 < t < T$, and then carrying the limit $T \to \infty$. The above method is particularly useful, however, because it leads directly not only to the manner in which the singularity at the origin is to be treated but also to the contour integral (976) and its evaluation. The use of these results in obtaining the corresponding response in linear passive networks will now be considered.

3. Response of linear systems. The problem of the communication engineer, formulated in general terms, is the following. Let the two-terminal pair N in Fig. 198 be any linear passive network, and Z_S and Z_R any linear passive impedances. With an arbitrary time function $e_1(t)$ impressed upon the input, we wish to know the corresponding output as characterized by either time function $i_2(t)$ or $e_2(t)$. Since we know that the function $e_1(t)$ may be considered as a linear superposition of sinusoidal (steady state) components in the Fourier sense, the problem is evidently solved in terms of the sinusoidal steady-state response alone. Thus if in particular we had

$$e_1' = E_1 e^{j\omega t}, \tag{977}$$

then in the steady state we would get

$$i_2' = I_2 e^{j\omega t}, \tag{977a}$$

where

$$I_2 = E_1 \cdot Y_{12}(\omega), \tag{977b}$$

in which $Y_{12}(\omega)$ is the complex transfer admittance function determined from the network parameters and the terminal impedances in the usual fashion.

When $e_1(t)$ is arbitrary we may, according to (966), represent it as

$$e_1(t) = \int_{-\infty}^{\infty} g(\omega) d\omega e^{j\omega t}. \tag{978}$$

Applying the principle of superposition we then have

$$i_2(t) = \int_{-\infty}^{\infty} g(\omega) \cdot Y_{12}(\omega) \cdot d\omega \cdot e^{j\omega t} \tag{978a}$$

as the desired response. The interesting and useful point in this result (other than its formal simplicity) is the fact that *the network response is completely specified by its sinusoidal steady-state behavior.* Since the integral (978*a*) extends over the entire (infinite) frequency spectrum, the steady-state response function $Y_{12}(\omega)$ must be known for *all* frequencies.

If instead of the current i_2 we wish to know the output voltage e_2 appearing across the load impedance Z_R, we get in an entirely similar fashion

$$e_2(t) = \int_{-\infty}^{\infty} g(\omega) \cdot h(\omega) \cdot d\omega \cdot e^{j\omega t}, \tag{978b}$$

in which

$$h(\omega) = Y_{12}(\omega) \cdot Z_R(\omega). \tag{979}$$

In order to formulate this situation more definitely, we may write the complex function $h(\omega)$ in its polar form

$$h(\omega) = H(\omega) \cdot e^{-j\theta(\omega)}, \tag{979a}$$

in which $H(\omega)$ is the magnitude, and $-\theta(\omega)$ the angle of $h(\omega)$. Here the angle is defined as $-\theta(\omega)$ in order that *lag* angles becomes *positive* and *lead* angles *negative.* The reason for this procedure will become clear in the following.

Using (968) and (979*a*) in (978*b*) we have

$$e_2(t) = \int_{-\infty}^{\infty} G(\omega) \cdot H(\omega) \cdot d\omega \cdot e^{j[\omega t + \psi - \theta]}. \tag{980}$$

Here $G(\omega) \cdot H(\omega)$ is evidently the *amplitude spectrum* of the output

voltage e_2, while $\psi(\omega) - \theta(\omega)$ is the corresponding *phase spectrum*. Thus we see that

> *The resulting amplitude spectrum is given by the* **product** *of the impressed and the response amplitude spectra, while the resulting phase spectrum is given by the* **difference** *between the impressed and the response phase spectra.*

It is now a simple matter to formulate the requirements of an *ideal* response system. Namely, if

$$\left.\begin{aligned} H(\omega) &= K = \text{constant} \\ \text{and} \qquad \theta(\omega) &= \omega t_d \end{aligned}\right\}, \tag{981}$$

where t_d is a constant, then (980) and (978), with the use of (968), show that

$$\begin{aligned} e_2(t) &= K \int_{-\infty}^{\infty} g(\omega) \cdot d\omega \cdot e^{j\omega(t-t_d)} \\ &= K \cdot e_1(t-t_d), \end{aligned} \tag{982}$$

which is an exact replica of $e_1(t)$ delayed by t_d seconds. The fact that a constant amplitude response and a linear phase response, as defined by (981), give rise to distortionless transmission, checks our former reasoning in this respect. We now see that the constant slope of the phase characteristic becomes the *time delay* of the system.

This formulation clearly distinguishes *two* kinds of distortion in a linear passive network. Namely, when the amplitude function $H(\omega)$ departs from a constant but $\theta(\omega)$ is a linear function, the network is said to possess **amplitude distortion** alone; when $H(\omega)$ is constant but the phase function $\theta(\omega)$ departs from linearity, the network is said to possess **phase distortion** (also called **delay distortion**) alone; when neither of the conditions (981) is fulfilled, the network possesses *both* amplitude and phase distortion.

Although the analytic formulation of the response in terms of the integral relations (978*a*), (978*b*), or (980) is a relatively simple matter, the evaluation of these integrals, except in very simple cases, is almost hopeless by ordinary means. However, for the purposes of the communication engineer, this situation is not at all a hopeless one. For, by studying a number of very simple examples, we are able to establish certain fundamental characteristics of network response which serve as criteria in estimating the sufficiency in the functions characterizing proposed networks for the transmission of intelligence in the form of a given time function. It is the object of the remaining sections of this

chapter to develop these criteria in terms of the specific treatment of a number of representative idealized cases.[1]

4. The response of idealized selective networks. In terms of the fundamental ideas regarding distortion in linear networks, as outlined above, the ideal filter may be thought of as a network having no phase distortion but a specific, preassigned type of amplitude distortion. Thus, in terms of the response functions (979*a*) (complex ratio of output-to-input voltage in the sinusoidal steady state), Fig. 199 illustrates the ideal characteristics of a low-pass filter. The amplitude function $H(\omega)$

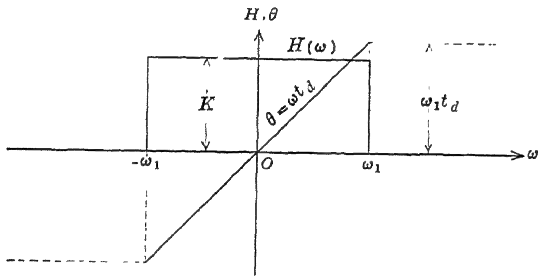


FIG. 199.—Amplitude (H) and phase (θ) response functions of the idealized low-pass filter.

is constant and equal to K only over the region $-\omega_1 < \omega < \omega_1$. Outside this range it is zero. The phase function is linear over the transmission range. Beyond this range the form of $\theta(\omega)$ is immaterial since nothing is transmitted anyway.

Let us assume various typical time functions impressed upon this idealized network and determine the resulting output functions. As the simplest type of impressed function let us consider the sudden application of a constant voltage E at $t = 0$, i.e., the function defined by (971) for $\alpha = 0$. We then have according to (973)

$$e_1(t) = \frac{E}{2\pi j}\int_{-\infty}^{\infty} \frac{e^{j\omega t}}{\omega}\,d\omega. \tag{983}$$

Since the amplitude spectrum for this function is

$$G(\omega) = \frac{E}{2\pi\omega}, \tag{983a}$$

we note that the vicinity of $\omega = 0$ is most important. We expect, there-

[1] The method of the following sections follows essentially that of K. Küpfmüller, "Über Einschwingvorgängen in Wellenfiltern," E.N.T., Vol. 1, 1924, pp. 141–152; also "Über Beziehungen zwischen Frequenzcharakteristiken und Ausgleichsvorgängen in linearen Systemen," E.N.T., Vol. 5, No. 1, 1928, p. 18.

fore, that the response characteristic of Fig. 199 will give rise to a fairly good reproduction of the impressed function.

Formulating the response according to (978*b*) and (979*a*), we have

$$e_2(t) = \frac{EK}{2\pi j}\int_{-\omega_1}^{\omega_1} \frac{e^{j\omega(t-t_d)}}{\omega}\,d\omega. \tag{984}$$

Replacing the exponential by its equivalent in terms of the sine and cosine functions, and using the abbreviation

$$\left.\begin{aligned} u &= \omega\,(t - t_d) \\ u_1 &= \omega_1(t - t_d) \end{aligned}\right\}, \tag{985}$$

this becomes

$$e_2(t) = \frac{EK}{2\pi j}\int_{-u_1}^{u_1} \frac{\cos u}{u}\,du + \frac{EK}{2\pi}\int_{-u_1}^{u_1} \frac{\sin u}{u}\,du. \tag{984a}$$

The first of these integrals has the same form as (975), and hence is given (except for the factor K) by (975*b*). For the evaluation of the

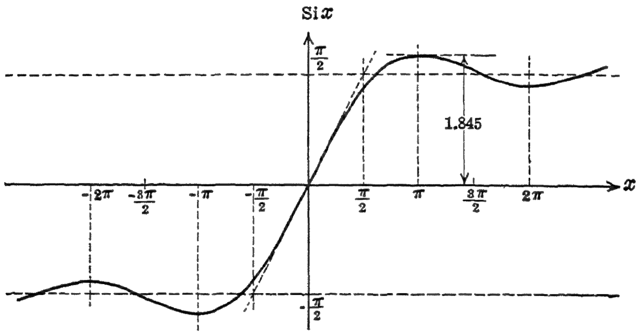


FIG. 200.—The function "sine-integral of x" defined by eq. (987).

second integral we note that the integrand is an even function so that we have

$$\int_{-u_1}^{u_1} \frac{\sin u}{u}\,du = 2\int_0^{u_1} \frac{\sin u}{u}\,du. \tag{986}$$

This last integral occurs so frequently in analytic work that tables of its values for various values of the upper limit (considered as an independent variable) are available.[1] This integral is usually written as

$$Si(x) = \int_0^x \frac{\sin u}{u}\,du, \tag{987}$$

and is called "the sine-integral of x." A plot of this function is shown in Fig. 200. It rises with a slope equal to unity at the origin and os-

[1] See for example Jahnke and Emde tables, p. 19 (first ed.)

cillates in a decaying manner about the value $\pi/2$, the maxima and minima lying at the zeros of the function $\sin u/u$, i.e., at multiples of π. Since the integrand is even, $Si(x)$ is an *odd* function of x.

The evaluation of (984*a*) is thus given by

$$e_2(t) = EK\left\{\frac{1}{2} + \frac{1}{\pi} Si\omega_1(t - t_d)\right\}. \tag{984b}$$

A plot of this is shown in Fig. 201. The time t_d, which we have called the delay-time, is seen to be the interval from $t = 0$ (the time at which E volts are suddenly applied at the input) to the time at which the output voltage has reached half its final value EK. If at this point we draw the tangent to the build-up curve (as indicated by the dotted line),

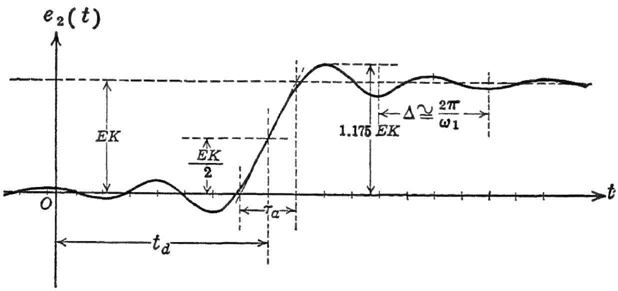


FIG. 201.—Response of the idealized low-pass filter, characterized by the functions of Fig. 199, when a constant voltage E is suddenly applied at $t = 0$.

then the time-interval between the intercepts of this tangent with the time-axis and the final value EK represents fairly well what may be called the **time of build-up.** This is denoted by τ_a in the figure, and is evidently equal to

$$\tau_a = \frac{EK}{\left(\dfrac{de_2}{dt}\right)_{t=t_d}} = \frac{\pi}{\omega_1}. \tag{988}$$

Thus, whereas the time delay is determined by the slope of the phase characteristic alone, the *time of build-up is inversely proportional to the width of the transmission band.* This statement, however, must be properly interpreted. Namely, the total phase shift throughout the pass band is determined by the extent of the structure (the least this can be in a physical network is π radians). Hence the slope t_d, if it is assumed constant, is inversely proportional to the cut-off frequency ω_1 (for a single constant-k section the average slope is π/ω_1). In this sense the time delay also depends upon the band width.

There is involved in this result for the function $e_2(t)$ another point

which may have caught the reader's attention upon closer examination of Fig. 201. This is the fact that the response is not zero for $t < 0$ as it obviously must be in a physical network. Since the mathematical analysis is not in error, we must conclude that the idealized amplitude and phase characteristics of Fig. 199 are not possible in a physical network. Another way of putting this is to say that these idealized characteristics are not compatible in a physical network; i.e., the amplitude and phase characteristics of physical networks are not independent but bear an implicit relation to one another which may be formulated explicitly by demanding zero response before $t = 0$ in the above problem. This point will be discussed in more detail in a following section.

For the present this point is of minor consequence since we are more particularly interested in the general character of the resulting behavior in this idealized case. Thus we note that, as the band width is increased, for a given total phase shift, both t_d and τ_a tend toward zero, and the resulting response, therefore, tends toward an exact replica of the form of the impressed force. A disturbing feature in this argument is due to the oscillatory character of the response. These oscillations, or *ripples* as we may call them, have a quasi-period equal to $2\pi/\omega_1$ seconds, and hence come closer together as ω_1 is increased. In the limit $\omega_1 \rightarrow \infty$ the ripple period becomes zero, but the response still surges above its ultimate value by the factor 1.175 for a differential instant. This peculiarity is in evidence whenever Fourier analysis is applied to functions with discontinuities and is known as the *Gibbs phenomenon*. Since the duration of this surge in the limit approaches zero, it may be ignored in the physical interpretation of this limiting case.

We conclude, therefore, that, since the applied disturbance contains all frequencies from zero to infinity, it can be accurately reproduced only by a system which faithfully transmits all these frequencies. An exclusion of part of the amplitude spectrum, even though there is no phase distortion, results in a distortion of the output relative to the input. In practice a certain amount of distortion can always be tolerated, and it is this fact which makes it possible to transmit signals by means of finite frequency bands. If it were not for this fact, intelligence communication as we know it today would be an impossibility. The region of frequencies which is of primary importance depends upon the nature of the applied signal. In our present example this is a suddenly applied constant voltage. For it, the region of largest frequency components lies in the vicinity of the origin. The low-pass filter may be expected, therefore, to do well in the way of transmitting this signal provided the cut-off is sufficiently high.

Suppose we vary the problem somewhat and assume that a sinusoidal emf with constant amplitude and a frequency ω_i somewhere within the pass band is suddenly applied to this same low-pass filter. According to (970) and (970a) we then have for $e_1(t)$, in place of (983)

$$e_1(t) = \Re e\left[\frac{E}{2\pi j}\int_{-\infty}^{\infty}\frac{e^{j\omega t}}{\omega - \omega_i}\,d\omega\right], \tag{989}$$

and for the output voltage $e_2(t)$ we have in place of (984)

$$e_2(t) = \Re e\left[\frac{EK}{2\pi j}\int_{-\omega_1}^{\omega_1}\frac{e^{j\omega(t-t_d)}}{\omega - \omega_i}\,d\omega\right]. \tag{990}$$

In order to facilitate the evaluation of this integral we replace ω by $\omega' + \omega_i$, taking account of the corresponding change in the limits of integration. After subsequently dropping the prime on ω we then have

$$e_2(t) = \Re e\left[\frac{EKe^{j\omega_i(t-t_d)}}{2\pi j}\int_{-(\omega_1+\omega_i)}^{\omega_1-\omega_i}\frac{e^{j\omega(t-t_d)}}{\omega}\,d\omega\right]. \tag{990a}$$

Writing the sine and cosine equivalent for the exponential and using for abbreviation

$$\left.\begin{aligned} u &= \omega(t - t_d) \\ u_1 &= (\omega_1 - \omega_i)(t - t_d) \\ u_2 &= (\omega_1 + \omega_i)(t - t_d) \end{aligned}\right\}, \tag{991}$$

this is

$$e_2(t) = \Re e\left[\frac{EKe^{j\omega_i(t-t_d)}}{2\pi j}\left\{\int_{-u_2}^{u_1}\frac{\cos u}{u}\,du + j\int_{-u_2}^{u_1}\frac{\sin u}{u}\,du\right\}\right]. \tag{990b}$$

For the second integral we have by (987)

$$\int_{-u_2}^{u_1}\frac{\sin u}{u}\,du = Si\ u_2 + Si\ u_1. \tag{992}$$

The first integral is more effectively divided as follows:

$$\int_{-u_2}^{u_1} = \int_{-u_2}^{u_2} - \int_{u_1}^{u_2}.$$

For the first of these we again have

$$\int_{-u_2}^{u_2}\frac{\cos u}{u}\,du = j\pi, \tag{992a}$$

so that

$$\int_{-u_2}^{u_1}\frac{\cos u}{u}\,du = j\pi - \int_{u_1}^{u_2}\frac{\cos u}{u}\,du. \tag{992b}$$

Making use of the function called the **cosine-integral**

$$Ci(x) = -\int_x^\infty \frac{\cos u}{u}\,du, \tag{992c}$$

this is equal to

$$j\pi + Ci\,u_2 - Ci\,u_1. \tag{992d}$$

Thus if we write

$$\left.\begin{aligned} \varphi_1(u) &= \frac{1}{2} + \frac{1}{2\pi}(Si\,u_2 + Si\,u_1) \\ \varphi_2(u) &= \frac{1}{2\pi}(Ci\,u_2 - Ci\,u_1) \end{aligned}\right\}, \tag{993}$$

and

$$\psi = \tan^{-1}\left(\frac{\varphi_2}{\varphi_1}\right) - \omega_i t_d, \tag{993a}$$

the evaluation of (990*b*) finally gives

$$e_2(t) = EK\sqrt{\varphi_1^2 + \varphi_2^2}\cdot\cos(\omega_i t + \psi). \tag{990c}$$

This represents an oscillation whose frequency is that of the impressed force, with a phase which depends upon the functions (993), enclosed in the envelope function

$$\varphi = EK\sqrt{\varphi_1^2 + \varphi_2^2}. \tag{994}$$

This function plotted versus time, therefore, shows the manner in which the output builds up.

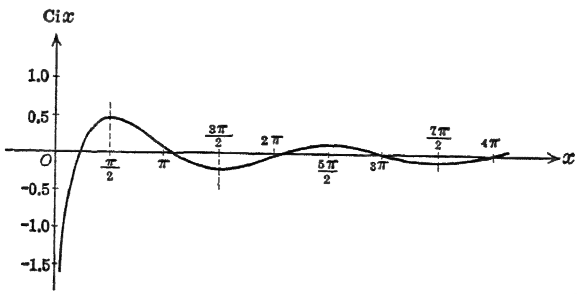


FIG. 202.—The function "cosine-integral of x" defined by eq. (992*c*).

A plot of the cosine integral (992*c*) is shown in Fig. 202 from which we see that its values are relatively unimportant except for small arguments. For these the following expansion is useful

$$Ci(x) = C + \ln x - \frac{x^2}{2\cdot 2!} + \frac{x^4}{4\cdot 4!} - \cdots,$$

in which $C = 0.5772$ is Euler's constant.

The envelope function (994) depends upon the behavior of φ_1 and φ_2 given by (993). These in turn depend upon the ratio of the impressed frequency ω_i to the cut-off frequency ω_1, according to the arguments u_1 and u_2 as expressed by (991). If $\omega_i = 0$, $u_1 = u_2$, so that $\varphi_2 = 0$ and

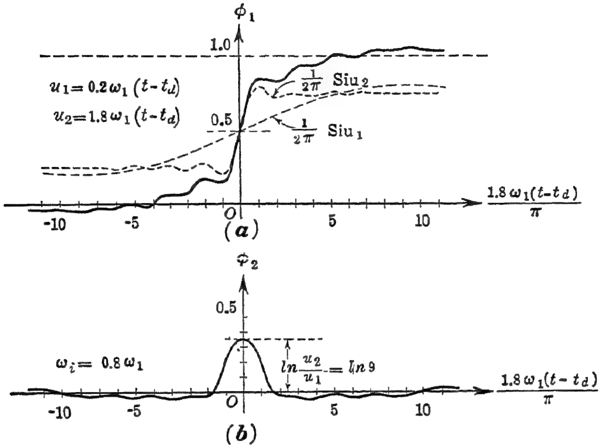


FIG. 203.—Component response functions (993) for the numerical case $\omega_i = 0.8\omega_1$.

$\varphi_1 = \dfrac{1}{2} + \dfrac{1}{\pi}\,Si\;\omega_1(t-t_d)$. Then φ is given by (984*b*), which is the result for the previous example. As an illustration of an intermediate case, suppose $\omega_i = 0.8\omega_1$. Then

$$\varphi_1 = \frac{1}{2} + \frac{1}{2\pi}\{Si\;1.8\omega_1(t - t_d) + Si\;0.2\omega_1(t - t_d)\},$$

$$\varphi_2 = \frac{1}{2\pi}\{Ci\;1.8\omega_1(t - t_d) - Ci\;0.2\omega_1(t - t_d)\}.$$

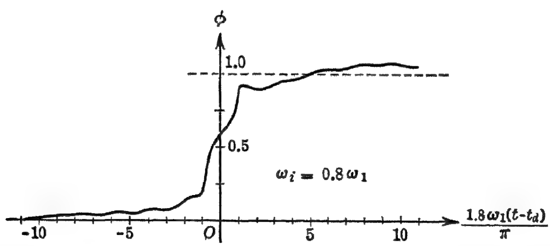


FIG. 204.—Response envelope (994) for the component functions of Fig. 203 which results when a sinusoidal voltage with frequency $\omega_i = 0.8\omega_1$ is suddenly impressed upon the idealized low-pass filter.

Plots of these are shown in parts (*a*) and (*b*), respectively, of Fig. 203, while the resulting φ is shown in Fig. 204.

We note from these figures that the effect of φ_2 upon the resulting envelope is confined to the immediate vicinity of the point $t = t_d$. At this point

$$\varphi_2 = \ln\left(\frac{\omega_1 + \omega_i}{\omega_1 - \omega_i}\right),$$

which is small unless ω_i is very close to ω_1. In that case the value of φ_2 at $t = t_d$ tends toward infinity, but it is doubtful whether the idealization of the network characteristics permits any conclusions to be drawn in such a limiting case.

Although it is quite evident that the time of build-up is longer than for the case $\omega_i = 0$, its formulation is not so clear. The tangent of the function φ at the point $t = t_d$ cannot be taken as determining τ_a in the manner in which this was done in Fig. 201. On the other hand, if we use the slope of φ_1 at $t = t_d$ we get the same value for τ_a as for the case $\omega_i = 0$. The slope of the function $\frac{1}{2\pi}$ $Si\ u_1$ seems to be a more reasonable criterion but perhaps somewhat pessimistic.

An interesting case occurs if we consider the impressed frequency to lie outside the pass band of the filter; i.e., $\omega_i > \omega_1$. Although the result for this may be evaluated from the previous treatment, the interpretation in terms of the Si and Ci functions is rather cumbersome. This may be avoided if we restrict our discussion to impressed frequencies whose value is several times the band width. Recognizing that the interpretation of results for frequencies near the cut-off is questionable anyway with the idealized characteristics, this is not a serious restriction.

Returning to the integral formulation (990) for the preceding case, we, therefore, proceed to evaluate on the assumption that $\omega - \omega_1 \cong - \omega_1$ for the region $-\omega_1 < \omega < \omega_1$, which is the same as replacing ω by its arithmetic mean value (zero). When $\omega_i \gg 2\omega_1$ this is a legitimate procedure. Assuming that the phase of the impressed frequency is α, i.e., $e_1 = E\cos(\omega_i t + \alpha)$ for $t > 0$, the integral (990) evaluates to

$$e_2(t) = -\frac{E\sin\alpha}{\pi}\cdot\frac{K\omega_1}{\omega_i}\cdot\frac{\sin\omega_1(t - t_d)}{\omega_1(t - t_d)}. \tag{995}$$

This is a pulse of the familiar form of $\sin u/u$ located symmetrically about the point $t = t_d$, with a maximum amplitude proportional to E and to $K\omega_1$, and inversely proportional to the impressed frequency ω_i. It is also interesting to note the dependence upon the phase α. Thus if $\alpha = 0$ the response is zero. This, of course, is only approximately correct on account of assumption made for the evaluation of the integral (990). A more exact evaluation would show a response for $\alpha = 0$

as well, although smaller than for $\alpha = \pi/2$. Since the impressed frequency lies outside the transmission range, the response is of a transient nature, as it should be.

Another interesting case in connection with the idealized low-pass filter of Fig. 199 is that for which we assume $e_1(t)$ to have the form of the Morse dot shown in Fig. 195, the Fourier integral for which is given by (965). Here the output voltage becomes

$$e_2(t) = \frac{\delta EK}{2\pi} \int_{-\omega_1}^{\omega_1} \frac{\sin \omega \frac{\delta}{2}}{\omega \frac{\delta}{2}} \cdot d\omega \cdot e^{j\omega(t-t_d)}. \tag{996}$$

Replacing the exponential by its sine and cosine equivalent we note that the resulting integrand for the term with $\sin \omega(t-t_d)$ is an odd function of ω. Since the limits are symmetrical about $\omega = 0$, the integral of this term vanishes. The other term involves

$$2 \sin \omega \frac{\delta}{2} \cos \omega(t - t_d) = \sin \omega\left(t - t_d + \frac{\delta}{2}\right) - \sin \omega\left(t - t_d - \frac{\delta}{2}\right).$$

Using the abbreviation

$$t' = \left(\frac{t - t_d}{\delta}\right), \tag{997}$$

and noting that the integrand is an even function, we have

$$e_2(t) = \frac{EK}{\pi} \int_0^{\omega_1} \frac{\sin \omega\delta(t' + \frac{1}{2}) - \sin \omega\delta(t' - \frac{1}{2})}{\omega} d\omega,$$

which evaluates to

$$e_2(t) = \frac{EK}{\pi} \{Si\, n\pi(t' + \tfrac{1}{2}) - Si\, n\pi(t' - \tfrac{1}{2})\}, \tag{996a}$$

with

$$\omega_1 = \frac{n\pi}{\delta}. \tag{997a}$$

It is now interesting to study this result for various values of the cutoff frequency ω_1. Reference to part (b) of Fig. 195 will show the convenience of the substitution (997a) for carrying this out. Fig. 205 shows plots of (996a) for the cases $n = 1, 2, 3, 4$, respectively. The dotted rectangle indicates the form of the input voltage and is added for purposes of comparison. The amplitude distortion of the idealized low-pass filter on the transmission of the Morse dot and its dependence upon the band width is made quite evident. The necessary band width for

which a random succession of dots of given duration will remain sufficiently distinct for good reception may readily be estimated on the basis of curves of this sort.

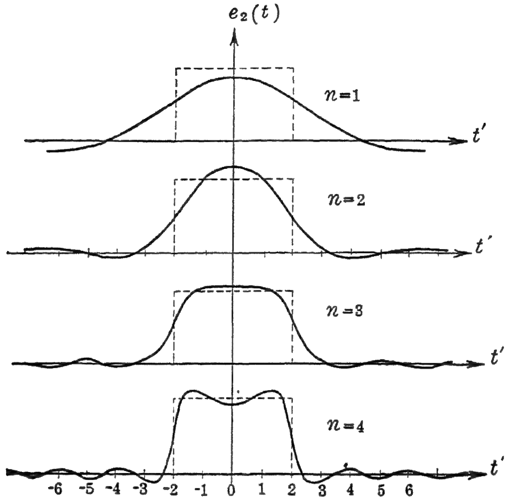


FIG. 205.—Response of the idealized low-pass filter for various values of its cut-off frequency $\omega_1 = n\pi/\delta$, when the Morse dot of Fig. 195 is the impressed voltage.

Let us now consider analogous examples for a band-pass filter. The idealized response characteristics for this case are drawn in Fig. 206. The cut-off frequencies are ω_{-1} and ω_1, respectively, and ω_0 is the mid-

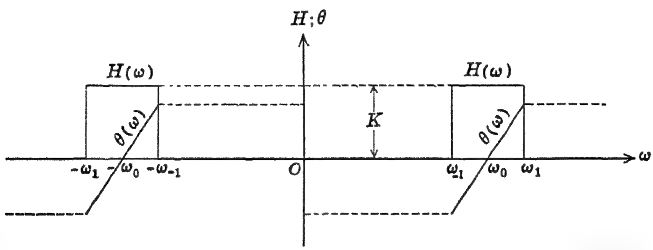


FIG. 206.—Amplitude (H) and phase (θ) response functions of the idealized band-pass filter.

band frequency. $H(\omega)$ and $\theta(\omega)$ are even and odd functions of ω, respectively. Within the pass band $\theta = (\omega - \omega_0)t_d$.

First we shall discuss the case for a suddenly applied sinusoidal emf with a frequency equal to ω_0. The Fourier integral for this is given

by (989) with ω_0 in place of ω_i. In formulating the response we note that the transmission range for negative frequencies will contribute relatively very little since the amplitude spectrum for the applied force is small in the region $-\omega_1$ to $-\omega_{-1}$ while it becomes infinite at $\omega = \omega_0$. Thus we shall restrict the integration to the positive band and get

$$e_2(t) = \mathcal{R}e\left[\frac{EKe^{j\omega_0 t_d}}{2\pi j}\int_{\omega_{-1}}^{\omega_1} \frac{e^{j\omega(t-t_d)}}{\omega - \omega_0} d\omega\right]. \tag{998}$$

Replacing ω by $\omega' + \omega_0$ and writing $w = \omega_1 - \omega_{-1}$, we have after dropping the prime

$$e_2(t) = \mathcal{R}e\left[\frac{EKe^{j\omega_0 t}}{2\pi j}\int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{e^{j\omega(t-t_d)}}{\omega} d\omega\right]. \tag{998a}$$

Applying similar principles to those used earlier, this evaluates to

$$e_2(t) = EK\left\{\frac{1}{2} + \frac{1}{\pi} Si\, \frac{w}{2}(t - t_d)\right\}\cos \omega_0 t, \tag{998b}$$

which is the same as the result for the low-pass filter with a constant voltage E impressed, as given by (984b), except for the factor $\cos \omega_0 t$. In fact, the latter follows from (998b) if we let $\omega_0 = 0$ (which makes $w/2 = \omega_1$). The plot of (998b) is shown in Fig. 207. The build-up time τ_a is obtained from the slope of the envelope function at the point $t = t_d$ and is

$$\tau_a = \frac{2\pi}{w}, \tag{988a}$$

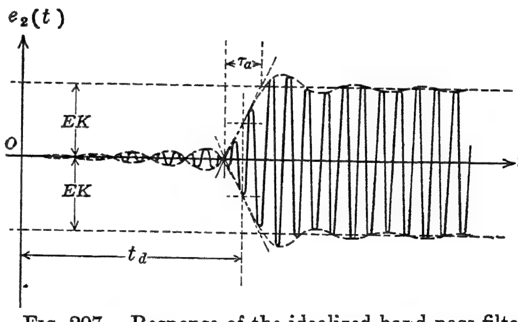


FIG. 207.—Response of the idealized band-pass filter when a sinusoidal voltage with frequency ω_0 (see Fig. 206) is suddenly impressed.

which is the same as (988) for the low-pass filter with $w/2$ in place of ω_1. Note that w is in radians per second. If w is expressed in cycles per second then $\tau_a = 1/w$. Thus a filter with a band width of 1000 cycles per second has a built-up time of 1/1000 second, etc. Incidentally, if the impressed emf should suddenly cease, the "build-down" time would be exactly the same. In fact, the time τ_a must be allowed for the output to respond to any abrupt change in the amplitude of the applied mid-band frequency.

This result clearly shows that a system which is to respond to an amplitude modulated carrier must pass not only the carrier frequency but also a band of frequencies whose width is determined by the rate at which the amplitude of the response is to follow corresponding amplitude variations in the carrier. Thus an infinitely selective system ($w \to 0$) does not respond to amplitude variations at all! Such a system has an infinite build-up time (besides having an infinite delay). The necessity for "side-band" frequencies is thus evident.

Recognizing the relation which this case bears to the response of the low-pass filter to a suddenly applied constant emf, the treatment for the band-pass filter with an applied frequency ω_i anywhere in the band is seen to be given by (990c) with (993) if in u_2 of equation (991) we replace $\omega_1 + \omega_i$ by $\omega_i - \omega_{-1}$. Fig. 204 then represents the envelope for the response of the band-pass as well as that of the low-pass filter.

When the applied frequency lies outside the pass band, the treatment for the band-pass filter differs somewhat from that for the low-pass primarily because in the band-pass the applied frequency may lie either above ($\omega_i > \omega_1$) or below ($0 < \omega_i < \omega_{-1}$) the band. Thus when ω_i lies below the band it may not be justified to neglect the contribution from the negative band as was done in the case just treated.

The response has the form of (998) in which ω_0 under the integral is replaced by ω_i, and the integration extends not only over the range $\omega_{-1} < \omega < \omega_1$ but over the range $-\omega_1 < \omega < -\omega_{-1}$ as well, recognizing that over the negative range $\theta = (\omega + \omega_0)t_d$ so that the sign of ω_0 in the exponential preceding the integral changes from plus to minus. Let us consider the contribution from the positive range first. If we again assume $|\omega_i - \omega_0| \gg w$, so that, for the integration, $\omega - \omega_i$ may be replaced by the constant average value $\omega_0 - \omega_i$, we have for this contribution

$$\mathcal{R}e\left[\frac{EKe^{j\omega_0 t_d}}{2\pi j(\omega_0-\omega_i)}\int_{\omega_{-1}}^{\omega_1} e^{j\omega(t-t_d)}\,d\omega\right] = \mathcal{R}e\left[\frac{EKe^{j\omega_0 t_d}}{2\pi j(\omega_0-\omega_i)}\cdot\frac{e^{j\omega_1(t-t_d)}-e^{j\omega_{-1}(t-t_d)}}{j(t-t_d)}\right] \tag{999}$$

But

$$e^{j\omega_1(t-t_d)} - e^{j\omega_{-1}(t-t_d)} = e^{j\frac{\omega_1+\omega_{-1}}{2}(t-t_d)}\left\{e^{j\frac{\omega_1-\omega_{-1}}{2}(t-t_d)} - e^{-j\frac{\omega_1-\omega_{-1}}{2}(t-t_d)}\right\}$$
$$= 2je^{j\omega_0(t-t_d)}\cdot\sin\frac{w}{2}(t-t_d), \tag{999a}$$

so that the contribution from the positive band becomes

$$e_2^+ = \frac{EKw}{2\pi(\omega_0-\omega_i)}\cdot\frac{\sin\frac{w}{2}(t-t_d)}{\frac{w}{2}(t-t_d)}\cdot\sin(\omega_0 t+\alpha), \tag{999b}$$

where α is the phase of the applied emf. The contribution from the negative band is obtained by replacing ω_0 by $-\omega_0$. The resulting response for this case, therefore, becomes

$$e_2(t) = \frac{EKw\mu}{\pi(\omega_0^2 - \omega_i^2)} \cdot \frac{\sin \frac{w}{2}(t - t_d)}{\frac{w}{2}(t - t_d)} \cdot \sin(\omega_0 t + \psi), \tag{1000}$$

where

$$\left.\begin{aligned} \mu &= \sqrt{\omega_0^2 \cos^2\alpha + \omega_i^2 \sin^2\alpha} \\ \psi &= \tan^{-1}\left(\frac{\omega_i}{\omega_0} \tan \alpha\right) \end{aligned}\right\}. \tag{1001}$$

When the impressed frequency lies above the band, the contribution from the negative band is relatively insignificant. Equation (999*b*) may

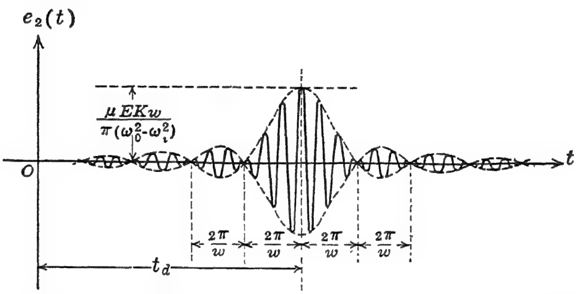


FIG. 208.—Response of the idealized band-pass filter when a sinusoidal voltage with frequency ω_i (lying outside the transmission band) is suddenly impressed.

then be considered as the entire response. This may still be true if $w/\omega_0 \ll 1$ and ω_i lies below the band but so that $(\omega_0 - \omega_i) \ll (\omega_0 + \omega_i)$.

A plot of the general case represented by (1000) is shown in Fig. 208. The maximum value of the transient oscillation is proportional to E and to Kw (which is the area under the amplitude characteristic of the filter), and inversely proportional to the difference between squares of the disturbing frequency and the mid-band frequency. The duration of the main impulse is $4\pi/w$ seconds, i.e., inversely proportional to the band width. It is also interesting to note that the frequency of the oscillation under the envelope is the *mid-band* frequency ω_0 and not that of the disturbing force. This result of course, is only approximately true but sufficiently correct so long as the disturbing frequency is far from the band compared to the band width. The result shows in an interesting manner how an undesired frequency may produce transient

interference within the transmission band of a filter when the amplitude of such a disturbing frequency changes abruptly.

5. Phase distortion. The above illustrations describing the transient behavior of idealized filters show the effect of amplitude distortion alone. In contrast to this we shall now consider the effect upon the transmission of a given signal, of departures from ideal behavior on the part of the phase function alone.

Regarding the output voltage (for example) as the function of interest, the amplitude and phase response functions according to (978*b*), (979), and (979*a*) are again given by $H(\omega)$ and $\theta(\omega)$. Here we shall assume $H(\omega)$ constant over the entire frequency range, but $\theta(\omega)$ arbitrary as illustrated in Fig. 209. The exact treatment for this case is considerably more difficult than for any of the preceding. The reasoning involved is the following: The entire spectrum may be thought of as divided into small increments $\Delta\omega = \omega_2 - \omega_1$ as illustrated in Fig. 209. If at the arithmetic mean frequency ω_0 of such an interval we draw a tangent to the θ-curve, then the phase characteristic may be taken approximately as the equation of this tangent so far as the contribution to the response from this frequency interval is concerned. This incremental response is thus similar to that of the idealized band-pass filter already treated, where ω_1 and ω_2 are substituted as the cut-off frequencies.[1] The net response is then

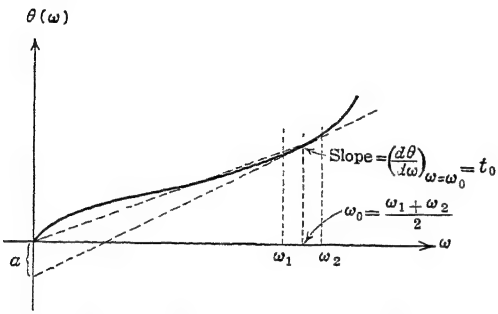


FIG. 209.—Arbitrary phase characteristic treated by dividing the frequency spectrum into small intervals $\Delta\omega = \omega_2 - \omega_1$ over which the slope is approximately constant.

[1] The fact that the phase characteristic for the increment $\Delta\omega$ does not pass through zero at ω_0 merely adds a constant phase angle to the corresponding incremental transient oscillation. This may be seen from the expression (1004) below for a wave group by setting x = const. (for the uniform line x = length of line). For the net response of the system to a given signal, the component responses for the incremental frequency groups must, of course, be summed (integrated) over the essential frequency range of the signal. Since the component constant phases differ for the various incremental responses, their integrated effect will in general affect the form of the net response. The exact treatment for a curved phase characteristic becomes very difficult, but as J. C. Steinberg points out (see "Effects of Phase Distortion on Telephone Quality," B. S. T. J. Vol. IX, pp. 550–556, 1930) the effect of constant phases due to the zero intercepts (called by him the *intercept distortion*) is relatively of subordinate importance as has been experimentally verified for the transmission

given by a summation of such incremental contributions over the whole frequency spectrum.

Since we are ordinarily not interested in the details of the response, it is unnecessary to carry out such a rather tedious summation. Instead, we may apply the following method for the approximate determination of the transient build-up time, which is the factor of chief interest. Among all possible tangents which may be drawn to the θ-curve, there is one which has a smaller slope than all the others. The contribution from the frequency increment located at this point evidently appears at the output terminals with the minimum delay, i.e., the time of arrival of this first contribution marks the practical *beginning* of the response. The time at which the response is practically built up depends upon the nature of the impressed disturbance. If the latter is a suddenly applied sinusoidal emf of frequency ω_i, then the frequency increment centered about ω_i involves by far the bulk of the energy contained in the disturbance.[1] The time of arrival of this major contribution, therefore, may be taken as marking roughly the *completion* of the transient build-up. Thus when phase distortion alone is involved, the transient time of build-up may be defined as

$$\tau_{ph} = \left(\frac{d\theta}{d\omega}\right)_{\omega=\omega_i} - \left(\frac{d\theta}{d\omega}\right)_{\min}, \tag{1002}$$

for the important case of a suddenly applied sinusoidal emf.[2] The *delay time* is here considered to be given by the *minimum* delay.

This method of viewing transient response involves the behavior of continuous frequency groups (increments) of small but finite width. It is appropriate, therefore, that we study the nature of such a group somewhat more in detail. For the sake of generality let us consider a wave group instead of a frequency group so that our results may be

---

of speech (see also J. C. Steinberg "Effects of Distortion on the Recognition of Speech Sounds," Jour. of the Acoustical Soc. of Am. Vol. I, pp. 121–137, Oct. 1929; C. E. Lane "Phase Distortion in Telephone Apparatus," B. S. T. J. Vol. IX, pp. 493–521, July 1930; and S. P. Mead "Phase Distortion and Phase Distortion Correction," B. S. T. J. Vol. VII, pp. 195–224). For a very comprehensive treatment of the problem of phase distortion the reader is referred to a series of three articles by H. G. Baerwald, "Ueber die Fortpflanzung von Signalen in dispergierenden Systemen," Ann. d. Phys. [5] 6, p. 295, 1930; [5] 7, p. 731, 1930; [5] 8, p. 565, 1931; here numerous additional references will also be found.

[1] This is evident from the fact that the amplitude spectrum for such an applied force reaches infinity at the frequency ω_i, but is otherwise finite.

[2] For other types of applied signal the time of completion for the transient build-up may be taken as the arrival time of the most essential frequency. In the case of the Morse dot, for example, this frequency is zero.

applied to systems having physical length (transmission lines) as well as to lumped networks.

The analytical representation of a differential wave component may evidently be written as

$$\mathcal{R}e[A(\omega) \cdot d\omega \cdot e^{j(\theta x - \omega t)}], \tag{1003}$$

in which θ is the phase function of the transmitting system and $A(\omega) \cdot d\omega$ the differential wave amplitude as a function of frequency. For any frequency interval such as that illustrated in Fig. 209, the phase function becomes $\theta = a + \omega t_0$, where t_0 is the slope of the tangent at $\omega = \omega_0$ and a its zero intercept. Over this small interval the amplitude function $A(\omega)$ may be considered constant and equal to, say, A_0. The continuous wave group corresponding to this finite frequency increment is then given by the integral

$$\mathcal{R}e \int_{\omega_1}^{\omega_2} A_0 e^{jax} \cdot e^{j\omega t_0 (x - t/t_0)} \cdot d\omega,$$

which evaluates to

$$A_0 \cdot \Delta\omega \cdot \frac{\sin\left[\frac{\Delta\omega}{2} t_0(x - v_g t)\right]}{\frac{\Delta\omega}{2} t_0(x - v_g t)} \cdot \cos \theta_0(x - v_0 t), \tag{1004}$$

where we have let

$$\left.\begin{aligned} v_g &= 1/t_0 \\ \theta_0 &= \omega_0 t_0 + a \\ v_0 &= \frac{\omega_0}{\theta_0} \end{aligned}\right\}. \tag{1004a}$$

Here θ_0 is evidently the value of θ at $\omega = \omega_0$, and v_0 the phase velocity at that frequency. The result (1004) represents a cosine wave with phase velocity v_0, contained within an envelope of the form $\sin u/u$ with a maximum amplitude $A_0 \cdot \Delta\omega$. The appearance of this envelope is the same as that given in Fig. 208. Its center is located at $x = 0$ for $t = 0$ and the length of the ripples is $2\pi/\Delta\omega$ instead of $2\pi/w$. This envelope evidently travels with a velocity v_g which is equal to the inverse slope of the θ-curve at the point $\omega = \omega_0$ and is referred to as the *group* or *envelope velocity.*

This result shows that a continuous wave group covering a frequency increment $\Delta\omega$ results in a single wave-impulse (except for some smaller ripples). A significant point is the fact that the velocity at which this

wave-impulse propagates is not the phase velocity (inverse slope of the line from the origin of the θ-plot to the curve at the frequency in question) but the group velocity (inverse slope of the curve at the frequency in question) which may be larger or smaller than the phase velocity.

These matters have to some extent already been discussed in connection with the long line in section 6 of Chapter III. We saw there that the group velocity of the long line (assuming constant parameters) is always larger than the phase velocity, and, over a large range of frequencies, larger even than $1/\sqrt{LC}$ which is the velocity of propagation of a wave front or discontinuity. The only conclusion to be drawn here is that those groups which theoretically arrive at the receiving end ahead of the front of the disturbance combine so as to cancel until the arrival of the wave front. The conclusion given above with regard to the build-up time does not apply in this peculiar case. In dealing with lumped networks where physical dimensions and hence physical propagation are not in evidence, this situation does not present itself so that the interpretation of the net response in terms of the superposition of incremental wave groups is always valid. Although a velocity of propagation is rather meaningless in connection with a lumped network, the reciprocal, namely, the time of arrival or response, does have a physical meaning.

The resultant form of a continuous frequency group as independently derived above could have been set down immediately on the basis of the discussion in section 1. Thus the functions illustrated in Fig. 195 (a) and (b) are each other's mates. Considering (a) as a frequency function (continuous frequency group) we have (b) as the corresponding time function. Applying the result expressed by (970) and (970a) we thus see that a pair of frequency groups located at $\pm\,\omega_0$ result in a time function consisting of a harmonic oscillation of frequency ω_0 enclosed in an envelope of the form of Fig. 195(b).

This analysis shows the general effect which certain departures from linearity on the part of the phase function have upon the transmission of a given form of signal. Thus a concave phase function (zero at the origin) causes more delay for the high frequencies than for the low ones. A Morse dot transmitted through such a system may be received as an elongated impulse similar to cases $n = 1$ or $n = 2$ of Fig. 205 followed by a high-frequency ripple, the severity of this type of distortion depending upon the concavity of the phase characteristic. In a system having a convex phase characteristic, on the other hand, the high-frequency ripple is received first and the main impulse later.

As an illustration consider a constant-k low-pass filter. It will be

recalled that the imaginary part of the propagation function per section is given in the transmission range by

$$\gamma_{2k} = 2 \sin^{-1}\left(\frac{\omega}{\omega_c}\right),$$

where ω_c is the cut-off frequency. Neglecting reflection effects this is the phase function for either voltage or current ratios. Assuming that this is a sufficiently good approximation for our present purpose, we have for the delay time[1]

$$t_d = \frac{d\gamma_{2k}}{d\omega} = \frac{2}{\omega_c\sqrt{1 - \left(\frac{\omega}{\omega_c}\right)^2}}, \tag{1005}$$

which shows that frequency groups approaching the cut-off are delayed considerably more than those in the vicinity of the origin, while the groups in the immediate vicinity of the cut-off are infinitely delayed. A frequency group at $\omega = 0.9\omega_c$, for example, is delayed 2.3 times as long as one at $\omega = 0$. When only a few sections of filter are involved, this is ordinarily not a serious matter since the delay at $\omega = 0$ (which is $2/\omega_c$ seconds per section) is so small that several times this value is still an unnoticeable amount in the transmission of speech or similar signals.[2] An interesting case where this variation in delay becomes important is that of a lump-loaded cable, which we have seen is to a first approximation the same as a low-pass constant-k filter of many sections. In a long circuit of this kind the minimum delay is appreciable, and the increase in delay for frequencies approaching the cut-off, therefore, is sufficient to produce noticeable squeals at the ends of spoken syllables, caused by the retardation of the higher frequencies. In transcontinental circuits, or wherever the minimum delay becomes comparable to the average duration of spoken syllables (or similar impulses), the requirement of linearity on the part of the phase function thus becomes important. The results of our present discussion serve as a guide in estimating allowable tolerances in this respect.

6. Simultaneous amplitude and phase distortion. The two previous sections dealt separately with the effects of amplitude and phase distortion upon the transient response of linear passive systems. This

[1] An alternative derivation of this result, based upon a more exact analysis is given on p. 18 of an article by Carson and Zobel entitled "Transient Oscillations in Electric Wave Filters," B.S.T.J., Vol. 2, No. 3, pp. 1–52. A general discussion of this subject is also found in "Electric Circuit Theory and Operational Calculus," by J. R. Carson, McGraw-Hill, 1926, pp. 117–131.

[2] For picture transmission the requirement on the phase characteristic is much more severe.

was done for the sake of simplicity and also for the purpose of emphasizing more clearly the individual effects. In practical systems, however, *both* amplitude and phase distortion are present to a greater or less extent depending upon the quality of the facility. The analytic evaluation of the response becomes very difficult here. We shall discuss only the general aspects of the problem. One of these has to do with the determination of the build-up time in the case of a suddenly applied sinusoidal emf of constant amplitude (the d-c problem is the special case for $\omega = 0$). Consider, for example, that the frequency of such an emf lies within the transmission band of a constant-k low-pass filter. If we assume, for the moment, that the distortion is due entirely to the non-linearity of the phase characteristic, then we have, according to (1002) and (1005),

$$\tau_{ph} = \frac{2}{\omega_c}\left\{\frac{1}{\sqrt{1 - \frac{\omega^2}{\omega_c^2}}} - 1\right\}. \qquad (1006)$$

A plot of this is shown in Fig. 210 in which the horizontal dotted line represents the build-up time for the idealized low-pass filter for the case of a suddenly applied constant emf. We recall that, when the applied emf for the idealized filter has a frequency somewhere within the band, the build-up time lengthens somewhat, particularly as the cut-off is approached. Although this does not strictly apply to the amplitude characteristic of the constant-k filter, we may conclude from Fig. 210 that over perhaps 0.8 of the band the effect of amplitude distortion considerably outweighs that due to phase distortion. Over this portion it is safe to consider the build-up time for the idealized low-pass filter (this may roughly be considered constant and equal to π/ω_c) as a sufficiently good approximation to the actual. For the range $0.8 < \omega/\omega_c < 1$ the effects of amplitude and phase distortion are of approximately the same magnitude so that no conclusions can be drawn on the basis of their separate effects.

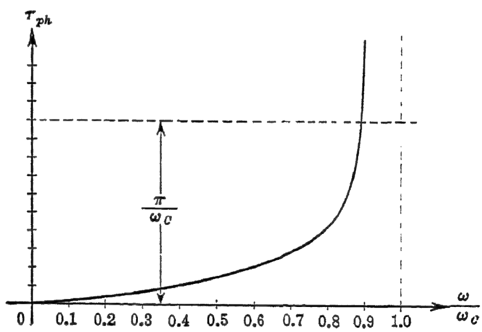


FIG. 210.—Time of build-up versus the frequency of a suddenly applied voltage, considering only the phase distortion in a constant-k low-pass filter. Dotted line is build-up time for the idealized low-pass filter with a constant voltage applied.

Thus, when amplitude distortion governs, the phase distortion may be neglected in determining the time of build-up, and vice versa. Conversely, we may argue that, if either $\tau_a \gg \tau_{ph}$ or $\tau_{ph} \gg \tau_a$, then either the amplitude or phase distortion, respectively, predominates. Cases which do not fall in one or the other of these alternatives cannot be treated even approximately by means of the simple methods outlined above.[1]

Another difficulty arises in connection with some practical circuits in attempting to apply the approximate relations (988) or (988*a*) for the time of build-up on the basis of amplitude distortion alone. This occurs where the rate of cut-off is so gradual that it becomes difficult to specify the band width with any degree of accuracy. Küpfmüller points out[2] that a reasonable value for the band width is obtained by dividing the total area under the $H(\omega)$-curve by the maximum ordinate usually occurring at the mid-band frequency (zero frequency in a low-pass filter—here the cut-off is half the band width).

A rather interesting case in point is the single-mesh R, L, C-circuit discussed in Chapter III of Volume I. The phase and amplitude characteristics for this network are shown in Figs. 49 and 50 (pp. 120 and 121, Vol. I). Since according to equation (269), page 118 (Vol. I), the area under the amplitude characteristic is infinite, the Küpfmüller rule gives rise to an infinite band width, which seems entirely unreasonable because we would thus have $\tau_a = 0$; i.e., we would ascribe no amplitude distortion to this circuit. On the other hand, if we determine τ_{ph} according to (1002), we find by means of equations (262) and (270), pages 118 and 120 of Volume I, for a suddenly applied mid-band frequency,

$$\tau_{ph} = \frac{2L}{R}.$$

Turning to equation (242*a*) and Fig. 39 (p. 100, Vol. I), we note that this value checks exactly with what we get by calculating that time at which a tangent to the envelope at the origin intersects the final value E/R. If we again apply (1002) to this circuit but for an arbitrary frequency we find

$$\tau_{ph} = \frac{2L}{R} \cdot \frac{1}{1 + \left(\frac{\omega - g}{\alpha}\right)^2},$$

where g is the resonant frequency and $\alpha = R/2L$. The actual build-up for the near-resonance case is given by equation (243) and is illustrated

[1] For a more exact treatment of certain special cases see J. R. Carson, "The Building-up of Sinusoidal Currents in Long Periodically Loaded Lines," B.S.T.J., Vol. 3, No. 4, 1924, pp. 558–566.

[2] K. Küpfmüller, *loc. cit.*, p. 31.

in Fig. 40a (pp. 100 and 101, Vol. I). Using the point of intersection of the tangent to the envelope at the origin with the final value as the time of build-up, we get for this case

$$\tau_{\text{actual}} = \frac{2L}{R} \cdot \frac{1}{\sqrt{1 + \left(\frac{\omega - g}{\alpha}\right)^2}}.$$

Although this does not exactly check τ_{ph} above, it is remarkably close to this value considering the approximations involved in the derivation of (1002).

It is significant to note that both τ_{ph} and τ_{actual} indicate that the build-up time for the off-resonance case is *less* than for the resonance case, whereas the reverse would be indicated on the assumption that amplitude distortion was the governing factor. Incidentally, if we apply (988a) p. 487 with $w = R/L$, which is usually taken as the "band width" of this circuit, we get $\tau_a = \pi \cdot 2L/R$ for the resonance case. This is evidently too large a value; i.e., assuming that amplitude distortion does govern the response, the effective band width ought to be taken about three times as large as is usually done in practice. We conclude, contrary to the common viewpoint, that the response of the series R, L, C-circuit is governed by its phase and not by its amplitude versus frequency characteristic. This illustration is interesting in that it shows the important part which the phase characteristic may play in network response.

Another interesting and important case of simultaneous amplitude and phase distortion results for a response function of the form

$$H \cdot e^{-j\theta} = \frac{ae^{-j\omega t_0}}{1 - be^{-j\omega t_1}}, \tag{1007}$$

in which a, b, t_0, and t_1 are constants. The amplitude and phase functions separately are given by

$$\left.\begin{aligned} H &= \frac{a}{\sqrt{1 + b^2 - 2b \cos \omega t_1}} \\ \theta &= \omega t_0 + \tan^{-1}\left(\frac{b \sin \omega t_1}{1 - b \cos \omega t_1}\right) \end{aligned}\right\}, \tag{1007a}$$

or if $b \ll 1$, approximately by

$$\left.\begin{aligned} H &= a(1 + b \cos \omega t_1) \\ \theta &= \omega t_0 + b \sin \omega t_1 \end{aligned}\right\}. \tag{1007b}$$

These are illustrated in Fig. 211.

In order to be able to recognize the type of response which results from this function we expand (1007) as follows on the assumption that $b < 1$.

$$H \cdot e^{-j\theta} = ae^{-j\omega t_0}(1 + be^{-j\omega t_1} + b^2e^{-2j\omega t_1} + \cdots)$$
$$= ae^{-j\omega t_0} + abe^{-j\omega(t_0+t_1)} + ab^2e^{-j\omega(t_0+2t_1)} + \cdots. \quad (1007c)$$

Each term in this series represents a response function for an ideal (distortionless) system, the first having a delay of t_0 seconds, the second a delay of $(t_0 + t_1)$ seconds with an amplitude factor b times that of the first, the third a delay of $(t_0 + 2t_1)$ seconds and an amplitude factor b^2 times that of the first, etc. Since b is assumed less than unity, the magnitudes of the successive responses (which are identical in form) diminish. The system evidently is one giving rise to *echo effects*, the relative magnitudes of the echoes being controlled by the factor b and the echo delay being given by t_1, while a and t_0 are the amplitude factor and time delay, respectively, for the initial response. The reader should note that, for the non-dissipative line terminated in pure resistances, the various ratios of output-to-input quantities (see section 17, Chapter II, pp. 71–76) have the form (1007) when the line is mismatched at its terminals. In such a case echo effects are evidently to be expected.[1]

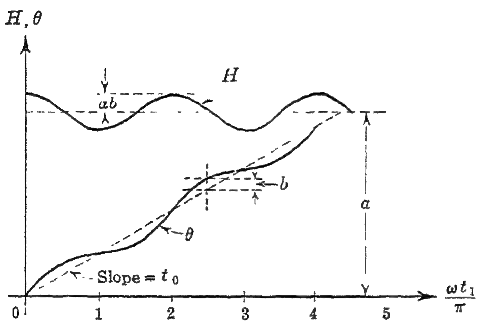


FIG. 211.—Amplitude and phase characteristics of a transmission network possessing echo effects, which are thus interpreted as a form of simultaneous amplitude and phase distortion.

7. The contour integral method of treatment. In section 2 above we pointed out that the Fourier integral for an arbitrary time function can be evaluated alternatively by applying the principles of complex integration. This is also true for the integrals (978*a*) and (978*b*) representing the current or voltage response of linear passive networks. This interesting fact, which leads to an important theorem regarding alternative methods of specifying network response functions, we wish now to elaborate upon in more detail. For this purpose let us apply the current response

[1] The method of expansion given here forms a basis for treating the general problem of the transient behavior of lines with arbitrary terminal networks in terms of the wave solution.

integral (987a) to the problem involving a suddenly applied sinusoidal emf of frequency ω_i. Utilizing the form (989) for the impressed emf, we have

$$i_2(t) = \mathcal{R}e\left[\frac{E}{2\pi j}\int_{-\infty}^{\infty} \frac{Y_{12}(\omega)\cdot e^{j\omega t}}{\omega - \omega_i}\,d\omega\right], \tag{1008}$$

or letting $\lambda = j\omega$, this is

$$i_2(t) = \mathcal{R}e\left[\frac{E}{2\pi j}\int_{-j\infty}^{j\infty} \frac{Y_{12}(\lambda)\cdot e^{\lambda t}}{\lambda - \lambda_i}\,d\lambda\right]. \tag{1008a}$$

For the discussion of the complex method, the form (1008a) is more convenient. We note that the integrand has a pole on the imaginary axis at the point $\lambda = \lambda_i = j\omega_i$ which must be avoided in the manner discussed in section 2. This is necessary whether we use the methods of complex integration or not. The remaining poles of the integrand are those of the admittance function $Y_{12}(\lambda)$. Following the method of defining closed contours for $t \gtrless 0$ as illustrated in Fig. 197 (except that the path is bulged at $\omega = \omega_i$ instead of at $\omega = 0$) it is clear that the response for $t < 0$ will be zero only if the poles of $Y_{12}(\lambda)$ all lie in the left-half λ-plane. Thus in a passive network this is evidently a physical requirement. This result is not new to us, however, since we know that the roots of the determinantal equation

$$Z_{12}(\lambda) = \frac{1}{Y_{12}(\lambda)} = 0,$$

which are the poles of the admittance function, must have negative real parts in passive networks in order that the transient (force-free) response shall decay with time. The method of contour integration gives us an interesting alternative view of this situation.

Since the poles of $Y_{12}(\lambda)$ must lie either on the real axis or come in conjugate complex pairs, the paths of integration for $t \gtrless 0$ are determined as illustrated in Fig. 212 in which the radius R is taken sufficiently large to enclose all the poles of $Y_{12}(\lambda)$ and insure a negligible contribution from the return path as already discussed.[1] Assuming, for the moment, that all the poles of $Y_{12}(\lambda)$ are simple (non-coincident roots of the determinantal equation $Z_{12}(\lambda) = 0$), a partial fraction expansion of the integrand takes the form

$$\frac{Y_{12}(\lambda)\cdot e^{\lambda t}}{\lambda - \lambda_i} = \frac{k_i}{\lambda - \lambda_i} + \frac{k_1}{\lambda - \lambda_1} + \frac{k_2}{\lambda - \lambda_2} + \cdots + \frac{k_n}{\lambda - \lambda_n}, \tag{1009}$$

[1] In order that R be finite this requires that $Y_{12}(\lambda)$ have no pole at infinity. The latter case which may occur in non-dissipative networks (by means of a pure capacitance branch) requires special treatment.

in which $\lambda_1, \lambda_2, \ldots \lambda_n$ are the poles of $Y_{12}(\lambda)$, i.e., the roots of the determinantal equation, and $k_1, k_2, \ldots k_n$ the residues of the integrand in these poles, while k_ι is the residue in the pole at $\lambda = j\omega_\iota$. Since these residues are given by

$$\left.\begin{aligned} k_\iota &= Y_{12}(\lambda_\iota)e^{\lambda_\iota t} = \frac{e^{j\omega_i t}}{Z_{12}(\omega_i)} \\ k_\nu &= \frac{e^{\lambda_\nu t}}{(\lambda_\nu - \lambda_i)\cdot Z_{12}'(\lambda_\nu)};\ (\nu = 1, 2, \cdots n) \end{aligned}\right\}, \tag{1010}$$

in which the prime denotes differentiation with respect to λ, the evaluation of (1008a) becomes

$$i_2(t) = \mathcal{R}e\left\{\frac{Ee^{j\omega_\iota t}}{Z_{12}(\omega_\iota)} + \sum_{\nu=1}^{n} \frac{Ee^{\lambda_\nu t}}{(\lambda_\nu - \lambda_\iota)Z_{12}'(\lambda_\nu)}\right\}. \tag{1011}$$

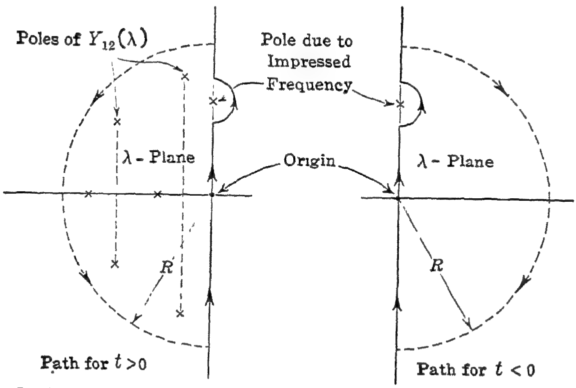


FIG. 212.—Paths of integration for the evaluation of the contour integral (1008a) showing the location of the poles of the transfer admittance function and of the impressed frequency.

For the special case of a suddenly applied constant emf $\omega_i = 0$. Here the bracket expression becomes real, so that we have

$$i_2(t) = \frac{E}{Z_{12}(0)} + \sum_{\nu=1}^{n} \frac{Ee^{\lambda_\nu t}}{\lambda_\nu \cdot Z_{12}'(\lambda_\nu)}. \tag{1011a}$$

It is interesting to note that this method of evaluation gives us the old familiar form of solution in terms of the steady-state plus the transient parts, these being represented by the first and second terms, respectively. The form of this result is that of Heaviside's expansion formula with which we are familiar from the discussion in Chapter IX of Volume I. This proof of the formula is due to K. W.

Wagner,[1] who also shows its extension to cases of coincident roots of the determinantal equation. For such the form of the partial fraction expansion (1009) and of the residues k_ν are evidently altered according to well-known principles of function theory which we shall not discuss in further detail here.

The following point is of interest to us at the moment. The two alternative methods of evaluating the Fourier integral, as discussed in this Chapter, show that the response of a linear passive network as characterized by the function $Y_{12}(\lambda)$ is determined *either by a knowledge of this function along the entire imaginary axis only, or by a knowledge of the location of the poles of* $Y_{12}(\lambda)$ *and of the residues of the function in these poles.*[2] Using the latter we arrive at the response in the more familiar form of steady-state plus transient parts, while, for the study of the behavior of idealized selective systems as discussed in section 4 above, the alternative specification of the response function in terms of its values along the imaginary axis is utilized.

8. The response functions of physical networks. In the discussion of the response of idealized selective networks we observe an obvious departure from physical behavior in that the output voltage is not zero before the time of application of the impressed voltage. Although the magnitude of this response is not large enough to be considered seriously in view of the primary purpose of these studies, nevertheless the reason for its appearance excites our curiosity. As we have already pointed out, this is due to the fact that the idealized amplitude and phase characteristics assumed for these networks are not physically possible, or that the pair of characteristics chosen are not physically compatible. Thus it appears that the amplitude and phase characteristics of physical networks *cannot be independently specified*; i.e., a certain choice for the amplitude characteristic (within physical limitations) restricts to a certain extent the phase characteristic which may be had in a physical network, and vice versa.

The Fourier integral method of treatment gives us a means for formulating this implicit relationship between the amplitude and phase characteristics of physical networks in a definite manner. This may be done in the following way: Suppose we assume a constant unit emf to be suddenly applied at $t = 0$ to the input terminals of a linear passive

[1] "Über eine Formel von Heaviside zur Berechnung von Einschaltvorgängen," Archiv für Elektrotechnik, Vol. 4, 1916, pp. 159–193.

[2] This is for the case of simple poles, i.e., non-coincident modes. If $\lambda = \lambda_\mu$ is a pole of multiplicity σ, then the residues of Y, $(\lambda - \lambda_\mu)\cdot Y$, . . . $(\lambda - \lambda_\mu)^{\sigma-1}\cdot Y$, in this pole must be known.

network with the response function (979a). According to the discussion in sections 2 and 3 we then have for the output voltage

$$e_2(t) = \frac{1}{2\pi j}\int_{-\infty}^{\infty} \frac{H(\omega)e^{j(\omega t-\theta)}}{\omega}\,d\omega, \tag{1012}$$

which is equivalent to

$$e_2(t) = \frac{1}{2\pi j}\int_{-\infty}^{\infty} \frac{H(\omega)\cos(\omega t-\theta)}{\omega}\,d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{H(\omega)\sin(\omega t-\theta)}{\omega}\,d\omega. \tag{1012a}$$

Consider, for the moment, the first of these integrals. Since $H(\omega)$ is always an even and $\theta(\omega)$ an odd function, it is clear that the integrand is an odd function of ω. Except for the fact that the integrand has a pole at the origin, the integral between symmetrical limits would vanish. The only contribution, therefore, comes from the immediate vicinity of the origin. For this vicinity we can replace the cosine by unity and $H(\omega)$ by $H(0)$, and have for this integral (ρ being a vanishingly small quantity)

$$\frac{H(0)}{2\pi j}\int_{-\rho}^{\rho} \frac{d\omega}{\omega} = \frac{H(0)}{2\pi j}\ln(\omega)\Big|_{-\rho}^{\rho} = \frac{H(0)}{2}. \tag{1013}$$

The algebraic sign is determined by the fact that the path of integration must pass slightly to the *right* of the origin as already discussed.

We thus have

$$e_2(t) = \frac{H(0)}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{H(\omega)\sin(\omega t-\theta)}{\omega}\,d\omega. \tag{1012b}$$

If the network is physical, this response must be zero for negative values of t, which is the same as to say that if we change the sign of t in (1012b) the result must vanish for $t > 0$. This gives

$$\frac{H(0)}{2} - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{H(\omega)\sin(\omega t+\theta)}{\omega}\,d\omega = 0 \text{ for } t > 0, \tag{1014}$$

which represents in integral form the relation sought. Thus only such amplitude and phase functions which satisfy (1014) may characterize a physical network.

This result is usually put in a slightly different form by differentiating with respect to t.[1] We then have

$$\int_{-\infty}^{\infty} H(\omega)\cdot\cos(\omega t+\theta)\cdot d\omega = 0 \text{ for } t > 0. \tag{1014a}$$

[1] The condition thus arrived at is less rigid because it virtually amounts to restricting the *derivative* of e_2 to be zero for $t < 0$, which allows e_2 to equal a constant before $t = 0$.

This may be put in a variety of alternative forms. For example, if we amplify (979a) to read

$$h(\omega) = h_1(\omega) + jh_2(\omega) = H \cos \theta - jH \sin \theta,$$

where $h_1(\omega)$ and $h_2(\omega)$ are the real and imaginary parts of $h(\omega)$, we have

$$\int_{-\infty}^{\infty} h_1(\omega) \cdot \cos \omega t \cdot d\omega = \dot{-} \int_{-\infty}^{\infty} h_2(\omega) \cdot \sin \omega t \cdot d\omega \text{ for } t > 0. \qquad (1014b)$$

Since $h_1(\omega)$ is even and $h_2(\omega)$ odd, this is equivalent to

$$\int_{0}^{\infty} h_1(\omega) \cdot \cos \omega t \cdot d\omega = - \int_{0}^{\infty} h_2(\omega) \cdot \sin \omega t \cdot d\omega \text{ for } t > 0. \qquad (1014c)$$

From these we may form explicit expressions of $h_1(\omega)$ in terms of $h_2(\omega)$ in the following way: According to the discussion of the Fourier transforms (966) we have for the case where $f(t)$ and $g(\omega)$ are both *even* functions

$$\left.\begin{aligned} f(t) &= \int_0^\infty g(\omega) \cos \omega t \, d\omega \\ g(\omega) &= \frac{2}{\pi}\int_0^\infty f(t) \cos \omega t \, dt \end{aligned}\right\}. \qquad (1015)$$

This is easily seen by replacing the exponential in (966) by its sine and cosine equivalent. The sine terms contribute nothing because the functions are even. For the cosine terms the limits may be changed from $-\infty$ to ∞, to 0 to ∞ by introducing a factor 2 and absorbing it in $g(\omega)$. Similarly if $f(t)$ and $g(\omega)$ are both *odd* functions we have the pair of sine-transforms

$$\left.\begin{aligned} f(t) &= \int_0^\infty g(\omega) \sin \omega t \, d\omega \\ g(\omega) &= \frac{2}{\pi}\int_0^\infty f(t) \sin \omega t \, dt \end{aligned}\right\} \qquad (1015a)$$

in which $g(\omega)$ is also a real function, the j having been absorbed in the manipulation from (966).

Returning to (1014c), we see that the cosine-transform of $h_1(\omega)$ equals the negative sine-transform of $h_2(\omega)$. Thus $h_1(\omega)$ is given by the cosine-transform of the negative sine-transform of $h_2(\omega)$, or

$$h_1(\omega) = -\frac{2}{\pi}\int_0^\infty \cos \omega t \, dt \cdot \int_0^\infty h_2(u) \sin ut \, du, \qquad (1016)$$

in which ω is replaced by u in the sine-transform in order to avoid con-

fusion of variables. Likewise $h_1(\omega)$ may be expressed as the negative sine-transform of the cosine-transform of $h_1(\omega)$, thus

$$h_2(\omega) = -\frac{2}{\pi}\int_0^\infty \sin \omega t \, dt \cdot \int_0^\infty h_1(u) \cos ut \, du. \qquad (1016a)$$

If the function $h_2(\omega)$ is specified, (1016) enables us to determine the corresponding $h_1(\omega)$. Conversely we may use (1016a) to determine $h_2(\omega)$ when $h_1(\omega)$ is given. The desired relations are thus formally set up. Unfortunately they are of little practical value, however, on account of the difficulty in evaluating the integrals even in the simplest cases.

Although these relations have been discussed here particularly with reference to the function $h(\omega)$, they apply generally to the real and imaginary parts of any physical impedance or admittance function. They also specify the relations between the magnitude and angle of such functions.[1] Thus if we write, for example,

$$h(\omega) = e^{-(A+jB)} \qquad (1017)$$

where $A(\omega)$ and $B(\omega)$ are the net attenuation and phase functions for the voltage ratio, then

$$\left.\begin{aligned} A(\omega) &= -\frac{2}{\pi}\int_0^\infty \cos \omega t \, dt \cdot \int_0^\infty B(u) \sin ut \, du \\ B(\omega) &= -\frac{2}{\pi}\int_0^\infty \sin \omega t \, dt \cdot \int_0^\infty A(u) \cos ut \, du \end{aligned}\right\}. \qquad (1016b)$$

Alternative integral formulations are given[2] for these relations but they are equally cumbersome to apply, and, therefore, we shall not consume space with them here. Our object is merely to emphasize more clearly that the pair of response functions of physical networks cannot be independently chosen.

9. Concluding remarks. It may seem to the reader that although the Fourier integral method of attack on the problem of network response is outwardly simple and compact, its usefulness in practice is not very great on account of the difficulty in evaluating, by analytic means, the integrals which are involved. If the object of studying this method were solely to obtain a process for the analysis of given physical cases, this would certainly be true, and in the opinion of the writer there would be little to recommend the inclusion of such a discussion here. To the communication engineer the Fourier integral method of reasoning and interpretation means much more than just an alternative proc-

[1] See Y. W. Lee, "Synthesis of Electric Networks by Means of the Fourier Transforms of Laguerre's Functions," Jour. of Math. & Phys., Vol. XI, No. 2, June, 1932, pp. 83–113.

[2] Lee, *loc. cit.*, p. 86.

ess of analysis of network behavior. In fact, in the great majority of cases, it is a means by which the actual labor of evaluating the response may be obviated entirely. It is this point of view that we wish to establish in the mind of the reader by our present remarks.

The strength of this process of thought rests largely upon the study of the response of idealized selective systems for certain types of impressed signals. Here, for example, the study of the response of the idealized low-pass filter with regard to an applied square-topped Morse dot of duration δ for various cut-off frequencies (illustrated in Fig. 205) may be useful in numerous ways. It shows that the faithfulness of reproduction varies directly with the duration of the impulse for a given width of transmission band. The shorter the impulse, the poorer will be the degree of reproduction. We recognize that, if we wish to transmit a complicated-looking signal with numerous irregularities and humps or indentations of short duration, the reproduction of the shortest of these, sets the criterion on the necessary band width for transmission. This is equally true if the signal is superposed upon a carrier frequency except that in the latter case the transmission band must be symmetrically located about the carrier frequency and its width is to be compared with twice the cut-off frequency of the low-pass problem for which the carrier is zero. The degree of reproduction having been picked from a series of idealized solutions such as illustrated in Fig. 205 (this will depend upon the practical requirements of the problem, of course), the necessary band width is readily determined. Thus we know that a facility with reasonably perfect amplitude and phase characteristics over this range will reproduce the applied signal with sufficient accuracy, and a subsequent analysis of the actual circuit for the actual signal applied becomes wholly unnecessary.

Similarly the study of phase or delay distortion becomes useful in establishing criteria for the maximum allowable deviation from linearity on the part of the overall phase function of a given facility. For example, if a facility is to be used for the transmission of speech, and we assume that the duration of the average syllable is, say, one-tenth of a second, then a variation in delay over the essential frequency band of that same order of magnitude will evidently produce sufficient garbling to make reception unintelligible. The allowable variation in delay will have to be kept considerably below that figure, the determination of this value depending upon a number of factors best disposed of by experiment. Nevertheless, once such a figure has been specified, the allowable variation (with respect to an average) in the slope of the phase characteristic is set, and the facility designed to meet this limit may be expected to perform satisfactorily.

Again in the idealized low-pass or band-pass filters, the fact that the build-up time for a suddenly applied mid-band frequency varies inversely as the band width clearly establishes a criterion for estimating the necessary band width when the speed with which the output is to follow amplitude variations on the input is set forth.

The Fourier method thus points out the requirements for ideal transmission and the rôles which are played by the amplitude and phase characteristics. It enables us to establish criteria for allowable departures from the ideal in terms of allowable departures in the faithfulness of reproduction. These are the cornerstones upon which methods of reasoning in the design and analysis of communication circuits rest.

This does not mean, of course, that exact analyses of communication network response are unessential. Such investigations have their collateral uses and are well worth the reader's study, although we shall not go into this matter here in more detail than that afforded by the Fourier method.[1]

[1] For a rather thorough study of the transient behavior of recurrent filter networks, the reader is referred to the article by Carson and Zobel given in the footnote on p. 494.

PROBLEMS TO CHAPTER XI

11-1. An idealized low-pass filter has a constant amplitude response function extending from zero up to an angular frequency of ω_1 radians per second. Its phase characteristic is given by the two confluent straight lines

$$\left.\begin{aligned}\theta &= \omega t_1;\ 0 \leq \omega \leq \omega_1 \\ \theta &= \omega t_2 + b;\ \omega_1 \leq \omega \leq \omega_2\end{aligned}\right\},$$

where $\omega_1 < \omega_2$ and $t_2 = t_1 - b/\omega_1$. The single Morse dot illustrated in Fig. 195 is impressed at the input. Show that the response does not remain finite unless b is an integer multiple of π. Assuming $t_1 = 5\delta$, determine and plot the response for

(a) $\omega_1 = \pi/\delta;\ \omega_2 = 3\pi/\delta;\ b = -\pi$,
(b) $\omega_1 = \pi/\delta;\ \omega_2 = 3\pi/\delta;\ b = \pi$,
(c) $\omega_1 = \pi/\delta;\ \omega_2 = 3\pi/\delta;\ b = -3\pi$,
(d) $\omega_1 = \pi/\delta;\ \omega_2 = 3\pi/\delta;\ b = 3\pi$.

11-2. A radio receiver is assumed to have at its input terminals an ideal band-pass filter with a band width of 10 kc. and is tuned to a distant station whose signal level at the receiver is -60 db. A local station whose signal level at the receiver is -20 db has its mid-band frequency located 40 kc. from that of the desired station. If the interfering carrier is suddenly applied or removed (worst condition) what will be the maximum approximate strength of the interference relative to the desired signal due to the transient effect?

11-3. A transmission system whose response characteristics may be approximated by those of the idealized low-pass filter as illustrated in Fig. 199 has transmitted over it a random succession of Morse dots of the form shown in Fig. 195. Consider only two such dots following each other at an interval of 1.25 δ seconds (center-to-center).

If the cut-off frequency is $\omega_1 = \pi/\delta; 2\pi/\delta; 3\pi/\delta; 4\pi/\delta$, determine and plot the resulting response. If the distinction between dots in the response is to be relatively clear cut, what should be the cut-off frequency of the given facility?

11–4. Determine the amplitude and phase spectra for the function $\cos pt$ whose amplitude is modulated by the Morse dot shown in Fig. 195. Show that, as δ is allowed to approach infinity, the amplitude spectrum approaches the single line at the angular frequency p.

11–5. Determine the amplitude and phase spectra of the Morse dot symmetrically located about the time $t = t_0$ instead of at the time origin as in Fig. 195.

11–6. For the spectrum of the amplitude modulated simple harmonic time function of Problem 11–4, show that the given time function is *not* reproduced by synthesizing only the upper or lower side bands except in the case $p = 0$.

11–7. Determine the amplitude and phase spectra for the time function defined by

$$\begin{aligned} f(t) &= 0;\ t < -\delta, \\ f(t) &= -E;\ -\delta < t < 0, \\ f(t) &= E;\ 0 < t < \delta, \\ f(t) &= 0;\ t > \delta. \end{aligned}$$

11–8. Consider a network whose amplitude response is constant between the frequencies $1/10\delta$ and $3/\delta$ cycles per second and whose phase response is linear, and determine the output for the applied signal defined in Problem 11–7.

11–9. Determine and plot the amplitude and phase spectra for the function defined by (971) for $\alpha = 0$. This function is impressed upon the idealized low-pass filter whose characteristics are illustrated in Fig. 199. Calculate and plot the response for the cut-off frequencies $\omega_1 = 2\pi \cdot 500;\ 2\pi \cdot 1000;\ 2\pi \cdot 5000$.

11–10. Determine the response for the function defined in the preceding problem when it is impressed upon the network whose characteristics are described in Problem 11–1.

11–11. Determine and plot the amplitude and phase spectra for the function $\cos pt$ whose amplitude is modulated by the function (971) for $\alpha = 0$.

11–12. A heavily loaded cable circuit may be approximately represented as a cascade of symmetrical low-pass constant-k sections for which the phase function is given by

$$\theta = 2n \sin^{-1} \frac{\omega}{\omega_c},$$

where n is the number of loading sections and ω_c the theoretical cut-off. If $\omega_c = 2\pi \cdot 10{,}000$ and $n = 88$, determine and sketch approximately the response for the impressed function described in Problem 11–4 with $p = 2\pi \cdot 1000$ and $\delta = 0.02$ second.

CHAPTER XII

SIMULATIVE AND CORRECTIVE NETWORKS

1. General remarks. The methods developed for the design of filters suggest the possibility of determining networks with prescribed attenuation, phase, or impedance functions of a more general character. Such networks are recognized as being of value as auxiliary aids in arriving at a prescribed degree of quality in the design of a communication facility. Thus, for example, the net propagation function of a transmission line or cable may be made to approximate its ideal more closely if properly cascaded with additional networks whose attenuation and phase characteristics very nearly counteract the existing discrepancies. Here the correction of amplitude and phase is usually thought of separately, and the auxiliary networks are then spoken of as **amplitude** and **phase corrective networks.** The former are also referred to as **attenuation equalizers.** The process, in general, is called **distortion correction.**

Another desirable function which may be accomplished by means of special networks is that of correcting the characteristic impedance of a line in order to make it more like a resistance over the essential frequency range. Thus a lump-loaded line, for example, may have a characteristic impedance differing considerably from its ideal. This situation may be improved by cascading a suitably designed network before terminating the line. This process is called **impedance correction,** and the network, an **impedance corrective network.**

Closely associated with this type of network is the so-called impedance **simulative network** or **line balance.** Its use becomes necessary wherever an image of the line impedance is called for, as, for example, in the duplex operation of telegraph circuits, or at repeater points in a two-wire circuit employing the same frequency channel for the two directions. In the latter case the degree of simulation obtainable definitely limits the gain at the repeater point.

Except for the phase corrective network, which is theoretically an all-pass non-dissipative structure, all these special networks involve dissipative elements. Our knowledge of network synthesis in the dissipative case is unfortunately relatively meager. The processes which have been developed for handling the problems mentioned above, even though they are able to meet practical situations fairly adequately, may still be referred to as being very decidedly in the "brute force"

stage. As such they are quite empirical and are characterized by design procedures whose details are difficult to formulate in a general manner. A presentation of the subject in its present stage, however, may give the reader a sufficient idea of the problem to stimulate further research.

2. Amplitude and phase distortion corrective networks. A thoroughly practical approach to this problem is given by Zobel,[1] who chooses primarily the symmetrical lattice as a theoretical point of departure. Recalling the discussion in section 1 of Chapter X, we note that this involves no loss in generality so far as the use of symmetrical structures is concerned. Assuming the desirability of a constant characteristic impedance, we begin by stipulating that the component impedances of the lattice shall satisfy the condition

$$Z_a Z_b = R^2 \tag{1018}$$

in which R is a positive real constant. Thus R becomes the constant value of the characteristic impedance Z_0. This assumes that the characteristic impedance of the network whose amplitude and phase functions are to be corrected (and to which the present structure is to be joined) also is a constant and equal to this same value. If this is so, then the overall propagation function equals the sum of the propagation functions for the given network and the cascaded corrective network.[2] The propagation function of the latter which will give a sufficient correction for the problem in hand is thus readily determined (most commonly, perhaps, in graphical form).

Recalling the first relation (846), page 379, for the symmetrical lattice, and making use of (1018), we get

$$\frac{Z_a}{R} = z_a = \tanh\frac{\gamma}{2} = \frac{e^\gamma - 1}{e^\gamma + 1}, \tag{1019}$$

or separating into real and imaginary parts by writing

$$\left.\begin{aligned} z_a &= r_a + jx_a \\ \gamma &= \gamma_1 + j\gamma_2 \end{aligned}\right\}, \tag{1020}$$

[1] O. J. Zobel, "Distortion Correction in Electrical Circuits," B.S.T.J., Vol. VII, pp. 438–534, July, 1928.

[2] When the structure to which the corrective network is to be joined does not have a sufficiently constant characteristic impedance, the loss and angle due to mismatch may be calculated by methods already discussed, and the required corrective characteristics determined so as to take account of these additional contributions to the net attenuation and phase functions of the given facility.

we find after some manipulation

$$\left.\begin{aligned} r_a &= \frac{\sinh \gamma_1}{\cosh \gamma_1 + \cos \gamma_2} \\ x_a &= \frac{\sin \gamma_2}{\cosh \gamma_1 + \cos \gamma_2} \end{aligned}\right\}, \tag{1019a}$$

from which r_a and x_a may be determined for given attenuation and phase functions.

In the design of corrective networks on this basis the thought is to determine, with the desired γ_1 and γ_2-functions, the corresponding real and imaginary parts of z_a. If a two-terminal network can then be found for this z_a-function, the problem is solved. Except in very special cases this procedure cannot be carried out because no physical network can be found for the realization of z_a. To begin with, we know from the discussion in section 8 of the previous chapter that γ_1 and γ_2 cannot be specified independently in a physical network, but must satisfy the relations (1016*b*). On the other hand, γ_1 or γ_2 alone does not specify the functions r_a and x_a according to (1019*a*), while the form of the relations (1016*b*) is such as to make their introduction into the present picture quite useless.

A graphical study of this situation not only affords a clearer view of these difficulties but also suggests a means for overcoming them. If, in the complex z_a-plane, we construct a family of loci for constant attenuation and another for constant phase, we may see quite readily the general character of the attenuation and phase functions obtainable from a given impedance z_a by sketching in this same plane the impedance locus for z_a. Our knowledge of certain characteristic types of impedance loci may thus suggest possible ways of meeting a specified network behavior, and make more evident the implied relation between the attenuation and phase functions.

The necessary analytic relationships for the construction of loci for constant γ_1 and γ_2 in the z_a-plane may be formulated in several ways. For example, if we eliminate either γ_1 or γ_2 from the relations (1019*a*) or from (1019) directly we find

$$\left.\begin{aligned} r_a^2 + x_a^2 - 2r_a \coth \gamma_1 &= -1 \\ r_a^2 + x_a^2 + 2x_a \cot \gamma_2 &= 1 \end{aligned}\right\}, \tag{1021}$$

which represent, in the z_a-plane, orthogonal families of circles for constant values of γ_1 and γ_2, respectively. For the circles on which γ_1 is constant the centers lie on the r_a-axis at the points $r_a = \coth\gamma_1$ with radii given by $1/\sinh\gamma_1$. The circles for constant values of γ_2 have

their centers on the x_a-axis at the points $x_a = -\cot\gamma_2$ with radii equal to $1/\sin\gamma_2$. This is illustrated in Fig. 213.

An alternative method of constructing these loci is based upon the inversion of (1019) directly, which reads

$$e^{\gamma} = \frac{1 + z_a}{1 - z_a}. \tag{1019b}$$

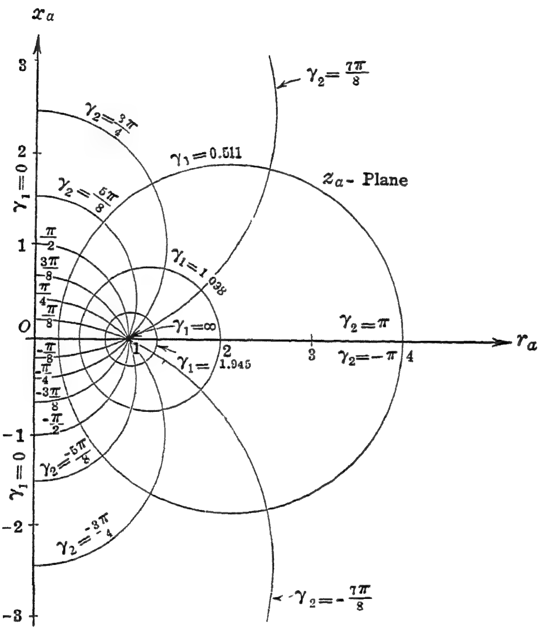


FIG. 213 —Loci of constant attenuation and phase according to the relations (1020) and (1021).

If we define

$$\rho_1 = |1 + z_a|\,;\ \rho_2 = |1 - z_a|$$

as the lengths of the vectors $(1 + z_a)$ and $(1 - z_a)$ respectively, then

$$\left.\begin{aligned} \gamma_1 &= \ln\left(\frac{\rho_1}{\rho_2}\right) \\ \text{and} \qquad \gamma_2 &= \measuredangle\,(1 + z_a) - \measuredangle\,(1 - z_a) \end{aligned}\right\}, \tag{1021a}$$

in which the sign $\measuredangle$ stands for the words "angle of." Fig. 213a shows the construction in the z_a-plane corresponding to these relations. The vector $(1 + z_a)$ is drawn from -1 to P, while $(1 - z_a)$ is given by the vector from P to $+1$. The angle γ_2 is thus seen to be that measured from the line $1,P$ to the extension of $-1,P$. This angle evidently remains constant if P moves along a circle for which $-1,1,P$ is the in

scribed triangle. The circle determined by this triangle, therefore, is the desired locus for constant γ_2. The ratio ρ_1/ρ_2, on the other hand, remains constant as P moves on a circle symmetrical with respect to the r_a-axis and passing through the points P and Q where PQ bisects the angle between ρ_1 and ρ_2 at P, thus dividing the length -1 to 1 in the ratio ρ_1 to ρ_2. The center of the circle for $\gamma_2 =$ constant is determined by the intersection of the perpendicular bisector of $-1,P$ with

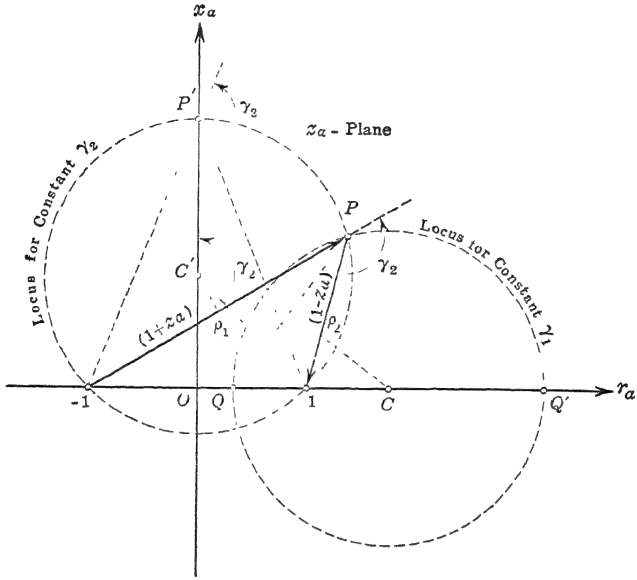


FIG. 213A.—The method of graphically constructing the loci of Fig. 213 according to the representation given by eqs. (1019b) and (1021a).

the x_a-axis at C', while that of the circle for $\gamma_1 =$ constant lies at the intersection of the perpendicular bisector of PQ with the r_a-axis at C.[1]

[1] The value of constant attenuation for a given circle is most readily calculated if we note that

$$\frac{\rho_1}{\rho_2} = \frac{1 + OQ}{1 - OQ} = \frac{OQ' + 1}{OQ' - 1}$$

where the point Q' is diametrically opposite Q and must evidently satisfy the relation $OQ \times OQ' = 1$. For a specified attenuation, OQ and OQ' and hence $OC = \frac{1}{2}(OQ+OQ')$ are thus easily obtained. The circle for a given constant angle γ_2 is most easily drawn if we note from the geometry that γ_2 is alternatively given by the angle $1\ C'\ P'$. The location of the center C' and the radius may, therefore, be immediately found by a simple construction. It should be noted in this connection, however, that for points on this circle lying below the r_a-axis, γ_2 differs by π radians from its value for points above this axis; i.e., the phase has a discontinuity of π radians as the point 1 is passed through.

It is interesting to note in passing that, if we regard this plane as being the cross-section through a pair of parallel circular-cylindrical conductors whose axes are located at the points -1 and $+1$ (like those of an ordinary uniform two-conductor line) carrying equal and opposite charges, then the loci for constant γ_2 represent the lines of electric field intensity and those for constant γ_1 the orthogonal equipotential surfaces. In our present problem, since a physical impedance must have a positive real part, only the right half-plane is to be considered. Here the attenuation values are evidently all positive as may be expected from a passive structure. The locus for zero attenuation is given by the x_a-axis and is realizable only by a non-dissipative network.

Several interesting points are immediately clear from an examination of the system of loci in Fig. 213. Since the latter are orthogonal we see that wherever the slope (with respect to frequency) of the attenuation function is zero (at a maximum or minimum), that of the phase function cannot be zero, and vice versa. Furthermore, at a large maximum in the attenuation function (close proximity to the point $+1$) the rate of change in the phase function is likely to be greater than at a smaller maximum, on account of the "bunching" of the lines of constant phase in the vicinity of the critical point $+1$. The latter, which corresponds to an infinite attenuation, is a singular point at which the phase changes abruptly by π radians.[1] Recalling that the real part of an impedance is an even function of frequency while the imaginary part is odd, we note that any impedance locus must be symmetrical with respect to the r_a-axis. Its tangent for zero or infinite frequency is perpendicular, therefore, to this axis. This makes it clear that the attenuation function must always have a maximum or minimum value at the origin and at infinity (i.e., it must be even) while the phase function at the origin must have a non-zero rate of change (i.e., it must be odd).

In general we note from Fig. 213 that as we traverse a closed contour in the clockwise direction so as to enclose the point $+1$, the phase increases continuously from zero (starting from a point on the r_a-axis between the origin and $+1$) to π for the first half of a revolution, and from π to 2π for the second half. A closed contour (closed impedance

[1] The value of the phase *in* the point 1 depends upon the direction from which this point is approached. In general, if the line of approach is at an angle of φ radians measured clockwise from O,1 then the phase becomes equal to φ as the point 1 is approached. In the case of a physical impedance the frequencies at which the real part is a maximum or minimum usually correspond very nearly (and in some cases exactly) to those frequencies at which the imaginary part passes through zero. This means that the impedance locus usually crosses the real axis very nearly (or exactly) at right angles. Hence if it passes through the point 1, the phase in this point then becomes $\pm\,\pi/2$.

locus) which lies wholly to the left or to the right of (or does not enclose) the point $+ 1$ will give rise to a phase function which rises and falls, or vice versa. Thus the general character of attenuation and phase functions obtainable from a given impedance locus becomes fairly evident; and, conversely, the general character of the necessary impedance locus for a desired pair of attenuation and phase functions as well as the manner in which they are related may be more easily pictured.

The Zobel method of procedure consists essentially of assuming a variety of networks for the impedance z_a and determining the corresponding general forms of attenuation and phase functions by means of the above or similar methods of analysis. In attempting to meet a specific design, that form of network for z_a is picked which most nearly gives rise to the form of attenuation function (usually this is more important) required and the parameter values in z_a determined from a sufficient number of equations expressing exact agreement between the desired attenuation function and that provided by this network at arbitrarily chosen frequencies. For simple attenuation functions this process is carried out fairly readily. In more complicated cases it is simpler to break the given attenuation function into component parts which may then more easily be realized by separate networks. The latter in cascade then yield the desired resultant attenuation function.

A difficulty which arises in more complicated cases is the fact that the process may not always lead to positive network elements; i.e., the physical realizability of z_a is not assured. Furthermore, it is not easy to see beforehand what kind of a phase characteristic the resulting network is going to have and whether its general shape will be such as to be helpful or otherwise. This latter point is not so essential, of course, since it is possible, by means of an all-pass structure, subsequently to correct for a remaining discrepancy in the phase without affecting the attenuation characteristic already obtained.[1]

Instead of giving a detailed account of the Zobel procedure here, we shall discuss an alternative method which, although not general, is sufficient to meet practical cases and has the advantage of being somewhat more flexible and lucid in the manner of its application. This method suggests itself if, in connection with the families of curves of Fig. 213, we recall the circle property of a familiar impedance locus.

[1] The phase characteristic obtainable from an all-pass network is restricted to having a *positive* slope, however. This fact is not a serious limitation (as will be seen later) provided the basic delay (resultant average slope) may be sufficiently increased to allow for a possible incremental corrective characteristic with negative slope. Thus an apparent negative slope requirement may be met by adding a linear characteristic with positive slope sufficient to make the resultant everywhere positive.

Suppose we choose for z_a the form of network shown in Fig. 214 in which r_0 and r are resistances and x represents a pure reactance. Analytically this impedance is given by

$$z_a = r_0 + \frac{rjx}{r + jx}, \tag{1022}$$

which may be written in the alternative form

$$z_a = r_0 + \frac{r}{2}(1 - e^{-j\theta}), \tag{1022a}$$

where θ is defined by

$$\left.\begin{aligned} \theta &= 2 \tan^{-1}\left(\frac{x}{r}\right) \\ x &= r \tan \frac{\theta}{2} \end{aligned}\right\}. \tag{1023}$$

The last relations show that, for a given value of r, θ is determined by the reactance x, and vice versa. Since a physical reactance is an increasing function with frequency, having alternate zeros and poles, θ increases with frequency, changing through 2π radians as x passes from a zero through a pole and back to zero, or from a pole through a zero back to a pole. The relation (1022a), therefore, shows that the locus of z_a as x varies is that of a circle of radius $r/2$, whose center lies on the real axis $r_0 + r/2$ units to the right of the origin. This is shown in Fig. 215. As x varies from zero through a pole and back to zero, for example, the tip of the z_a-vector makes one complete revolution starting from the point $r_a = r_0$, $x_a = 0$, which is the point from which θ is measured in the clockwise direction.

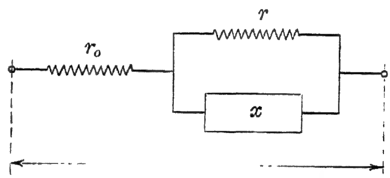


FIG. 214.—Impedance whose analytic form is expressed by the relations (1022), (1022a), and (1023).

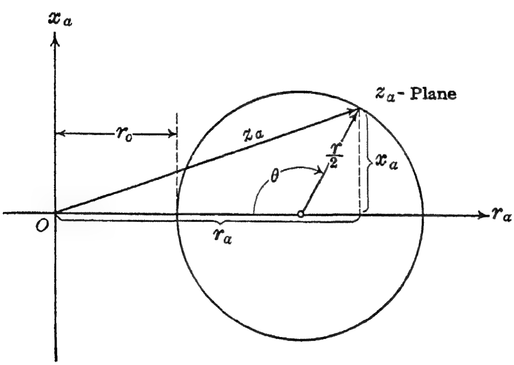


FIG. 215.—Locus in the z_a-plane of the impedance of Fig. 214.

This circle property of the impedance of Fig. 214 is well known in the study of impedance

loci. Since the reader probably is not familiar with this subject it may be well to review briefly a few of the basic ideas here. A fundamental geometrical proposition is illustrated in Fig. 216. The line through the points P and Q is parallel to the vertical reference axis. If P represents any point on this line, then the inverse point P' lies on the circle tangent to the vertical reference axis at the origin. The proof of this, which depends upon the fact that the triangles OPQ and $OQ'P'$ are similar, is easily given. We have

$$\frac{OP'}{OQ} = \frac{OQ'}{OP}.$$

If Q' is so located with respect to Q that

$$OQ \cdot OQ' = 1,$$

then

$$OP \cdot OP' = 1.$$

The circle is said to be the *inversion* of the straight line with respect to unity. Evidently, if the line lies to the right of the point 1 (as in Fig. 216), the circle lies to the left of this point; if the line lies to the left, then the circle encloses the point 1.

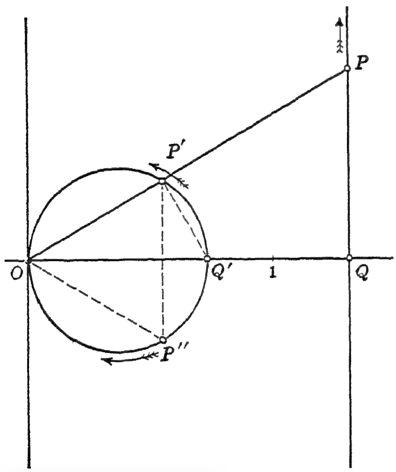


FIG. 216.—Geometrical construction illustrating the inversion of a straight line.

Conversely we see that the inversion of a circle tangent at the origin is a straight line. When the circle is not tangent but lies to the right of the origin, then its inversion is again a circle lying to the right of the origin. This is illustrated in Fig. 216a for the case where the given circle lies either wholly to the right or to the left of the point 1. When this point is enclosed, the circles overlap; i.e., both circles enclose the point 1. The proof is again based upon similar triangles. For example, since the triangles ONB and $OB'N'$ are similar we may write

$$\frac{ON}{OB} = \frac{OB'}{ON'},$$

and if

$$OB \cdot OB' = 1,$$

we have

$$ON \cdot ON' = 1.$$

Hence the points N and N' are inverse. Other corresponding inverse points are readily recognized from the figure. It is also significant to note the relative directions in which the points P and P' traverse the loci in Figs. 216 and 216*a*.

In the case of impedances, the planes in these figures are complex. This introduces only one additional feature, namely that the inverse of P is P'', the *image* of P' with respect to the real axis. Note that this reverses the direction in which the inverse locus is traversed. If the

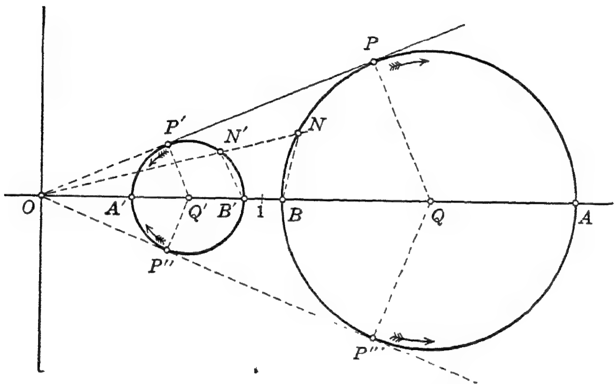


FIG. 216A.—Geometrical construction illustrating the inversion of a circle.

line of Fig. 216 is traversed from bottom to top, then P'' traverses the circle in the clockwise direction, and if P in Fig. 216*a* rotates clockwise, then P'' does likewise.

The straight-line locus in Fig. 216 is evidently that of an impedance consisting of a fixed resistance in series with a variable reactance. If we reciprocate this with respect to unity, the resulting impedance, which consists of the inverse resistance and reactance in parallel, has the circular locus tangent at the origin. This is the case for the impedance z_a of Fig. 214 for $r_0 = 0$. When $r_0 > 0$, the locus is a circle lying to the right of the origin as shown in Fig. 215. The impedance z_b in the *cross* arms of the lattice then has a circular locus determined according to Fig. 216*a*. It is also clear that the circular loci of z_a and z_b either lie to the right and left respectively of the point 1 (in inverse relationship) or both lie so as to enclose this point. Since interchanging the im-

pedances in the arms of a lattice merely has the effect of changing the phase by π radians (and this is of no consequence), we note the interesting fact that in the attenuation and phase chart of Fig. 213 the regions to the right of a vertical line through the point 1 and within a tangent circle through this point are *corresponding regions* in the sense that impedances (which may be chosen, say, for z_a) with loci within these give rise to potentially identical corrective characteristics. These regions are shown cross-hatched in Fig. 217. If we choose the form of network of Fig. 214 for z_a, we have to distinguish only two essentially different cases: Either the locus lies to the right or left of the singular point 1, or it lies so as to enclose this point.

From our previous discussion of the attenuation and phase chart of Fig. 213 we readily see the general character of functions obtainable when this type of impedance is used for z_a. As the circular locus is traversed, i.e., as the reactance x increases with increasing frequency, the attenuation varies between maximum and minimum values which depend upon the choice of r_0 and r only. The number of equal maxima and minima depend upon the number of zeros and poles of the reactance x within the prescribed frequency range. The detailed form of the attenuation function depends upon the detailed form of the reactance function and hence upon the number and location of its zeros and poles *outside* the prescribed frequency range. If the required attenuation function is to have a number of non-equal maxima or minima, the desired result can obviously not be obtained from a single lattice with component impedances of this type. By breaking the given attenuation characteristic into components and cascading lattices giving the individual components, however, practically any desired form of resultant characteristic may be obtained. For a component network, r_0 and r are determined by the extremal values within which the attenuation is to vary. With these resistances fixed, the detailed form of the attenuation function determines the necessary reactance x as a function of frequency. The problem is then reduced to that of finding a physical non-dissipative, two-terminal structure whose reactance function approximates x with sufficient accuracy to meet prescribed tolerances. Since simultaneous variations in r_0 and r and the

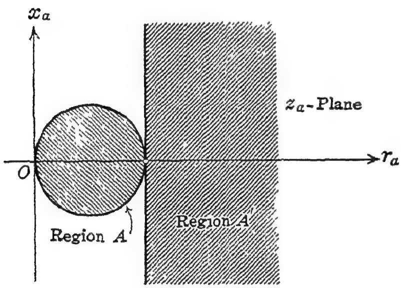


FIG. 217.—Regions corresponding to inverse impedance loci.

limits of the approximation range affect the resulting x-function, a number of trials may be required before an x-function realizable by means of elements with convenient magnitudes is obtained.[1]

The corresponding general form of phase characteristic which is simultaneously obtained by this process depends primarily upon whether the impedance locus is located so as to enclose the point 1 or not. In the former case the phase characteristic rises continuously; in the latter, it rises and falls, or vice versa. Between these two tendencies one has a clear choice, but no choice is left as to the detailed form once the attenuation function has been set. An interesting case results when the locus of z_a is so placed as to coincide with a circle for constant attenuation. Then the impedance locus encloses the point 1, and hence only a continuously increasing phase function may be had. The special case for zero attenuation leads to the non-dissipative all-pass network used for phase correction alone.

The detailed aspects of this method of corrective network design are more clearly seen if we transform our loci from the z_a-plane to the corresponding ones in the e^γ-plane. Using the expression (1022) for z_a in (1019b), and letting

$$\left.\begin{aligned} a &= \frac{1 + r_0}{1 - r_0} \\ b &= \frac{1 + (r_0 + r)}{1 - (r_0 + r)} \end{aligned}\right\}, \tag{1024}$$

we find

$$e^\gamma = \frac{b + a}{2} - \frac{b - a}{2} e^{-j\psi}, \tag{1025}$$

with

$$\psi = 2 \tan^{-1}\left(\frac{a + 1}{b + 1} \cdot \frac{x}{r}\right). \tag{1025a}$$

This shows that in the e^γ-plane the impedance locus for z_a is again a circle, being given by (1025) in which the angle ψ increases continuously with x according to (1025a) and hence increases continuously with frequency. This result is illustrated in Fig. 218.

In this plane the loci for constant attenuation are concentric circles and those for constant phase are radii. The region within the unit circle corresponds to negative attenuation and hence is not available for passive networks. It corresponds to the negative half of the z_a-plane. The right half of this plane is transformed into the region outside the

[1] This choice may also determine whether or not a physical network can be found for the realization of x. This question is further discussed on pages 526 and 527.

unit circle in the e^{γ}-plane of Fig. 218. The interesting point is the fact that the circular impedance locus in the z_a-plane again appears as a circle in the e^{γ}-plane. The symmetry of the loci for constant attenuation and phase in this plane makes it easier to see the various general relationships pointed out earlier with regard to possible results obtainable

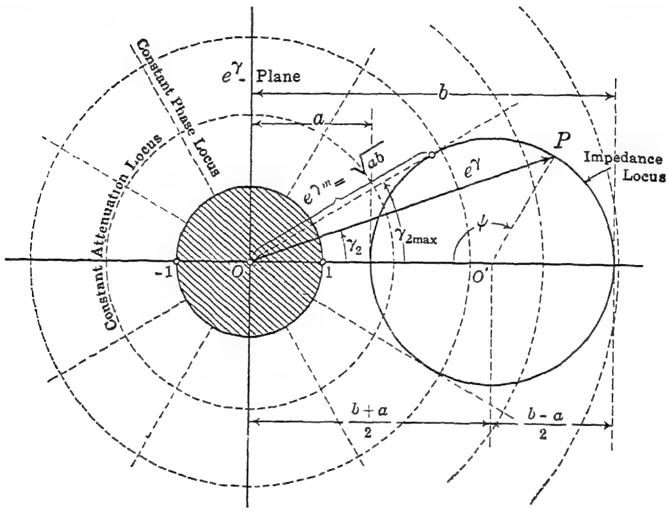


FIG. 218.—Conformal map of the impedance locus of Fig. 215 in the e^{γ}-plane.

from the circular locus of the impedance z_a. Thus if we set down the three possible cases occurring for relative values of r_0 and r, we have with reference to (1024)

Case I

$$\left.\begin{array}{c} 0 < r_0 < 1; 0 < (r_0 + r) < 1 \\ 1 < a < b \end{array}\right\},$$

Case II

$$\left.\begin{array}{c} 0 < r_0 < 1; 1 < (r_0 + r) < \infty \\ a > 1; b < -1 \end{array}\right\},$$

Case III

$$\left.\begin{array}{c} 1 < r_0 < \infty; 1 < (r_0 + r) < \infty \\ a < b < -1 \end{array}\right\}.$$

Here cases I and III evidently correspond to impedance loci within the regions A and A', respectively, as shown in Fig. 217 for the z_a-plane. In the e^{γ}-plane these correspond to the regions to the right of the point 1

and to the left of -1, respectively, as illustrated in Fig. 217*a*. The symmetry of the e^γ-plane makes it clear that the results obtainable from loci in these regions are potentially the same except for a difference of π radians in the phase. This was not so evident in the z_a-plane. It should also be noted from the relation (1025*a*) for ψ that in both cases I and III the point P (tip of e^γ-vector in Fig. 218) rotates clockwise with increasing frequency.

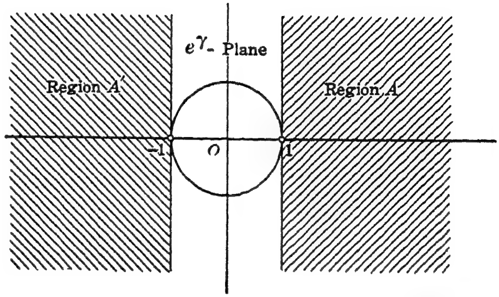


FIG. 217A.—Regions in the e^γ-plane corresponding to those in the z_a-plane shown in Fig. 217.

Case II is that for which the impedance locus encloses the point 1 in the z_a-plane or the unit circle in the e^γ-plane. The point P in Fig. 218 then rotates counter-clockwise so as to give an increasing phase function with frequency. Writing

$$-b = \frac{1 + \dfrac{1}{r_0 + r}}{1 - \dfrac{1}{r_0 + r}},$$

we note that

$$\left.\begin{array}{ll} & a \gtrless -b \\ \text{according to} & r_0(r_0 + r) \gtrless 1, \text{ or } r \gtrless \dfrac{1 - r_0^2}{r_0} \end{array}\right\}. \qquad (1024a)$$

Thus for $r > \dfrac{1 - r_0^2}{r_0}$, $|a| > |b|$, so that the center of the impedance locus in Fig. 218 lies to the right of the origin, while for smaller values of r the center lies to the left. In the intermediate case defined by the equality sign, the impedance locus is concentric with the origin and a constant attenuation results.

In order to study the detailed relations between the attenuation and phase as functions of the arbitrary reactance x (refer again to Fig. 214 for the impedance z_a), it is effective to introduce the notation

$$\left.\begin{array}{l} a = e^{\gamma_a};\ \gamma_a = \ln a \\ b = e^{\gamma_b};\ \gamma_b = \ln b \end{array}\right\}. \qquad (1026)$$

Since a and b mark the limits of the impedance locus in Fig. 218, we see that γ_a and γ_b are the extremal values within which the attenuation

varies. In case a or b is *negative*, γ_a or γ_b has an imaginary part equal to $\pm j\pi$. The real part then signifies the extremum in the attenuation.

It is further convenient to define as the *attenuation interval*

$$\delta = \gamma_b - \gamma_a = \ln\left(\frac{b}{a}\right), \tag{1027}$$

and as the *arithmetic mean attenuation* within this interval

$$\gamma_m = \frac{\gamma_b + \gamma_a}{2} = \ln\sqrt{ab}. \tag{1028}$$

Then if we let

$$y = \left(\frac{\cosh\frac{\gamma_a}{2}}{\cosh\frac{\gamma_b}{2}}\right)\cdot\left(\frac{x}{r}\right), \tag{1029}$$

which, for a given choice ot a and b is a frequency function directly proportional to the reactance x, we find, after some manipulation of (1025) and (1025a), the following relation between attenuation γ_1 and the proportional reactance y

$$e^{(\gamma_1 - \gamma_m)} = \sqrt{\frac{a + by^2}{b + ay^2}} = \sqrt{\frac{1 + e^{\delta}y^2}{e^{\delta} + y^2}}, \tag{1030}$$

or the inverse relations

$$\left.\begin{aligned} y &= \sqrt{\frac{b}{a}}\cdot\sqrt{\frac{e^{2\gamma_1} - a^2}{b^2 - e^{2\gamma_1}}} = \sqrt{\frac{\sinh(\gamma_1 - \gamma_a)}{\sinh(\gamma_b - \gamma_1)}} \\ &= \sqrt{\frac{\sinh[\delta/2 + (\gamma_1 - \gamma_m)]}{\sinh[\delta/2 - (\gamma_1 - \gamma_m)]}} \end{aligned}\right\}. \tag{1030a}$$

The significant point of this result is the fact that the proportional reactance function y *depends only upon the attenuation interval* δ for a specified functional variation in $(\gamma_1 - \gamma_m)$ versus frequency. Thus the required y-function in a given case is independent of the average attenuation level in the prescribed attenuation function. This means that, in a given design, the reactance x may be determined to within a constant multiplier without for the moment having to consider the average attenuation level of the required function. This is a decided convenience as will be more fully discussed below. Since the design may thus be approached without initially choosing specific values for a and b (or the attenuation limits) we recognize as an added flexibility in the method that the location of the impedance locus in Fig. 218 is left arbitrary during the process of meeting prescribed attenuation requirements. Having finished this part of the design, we may choose between cases

I, II, or III above; i.e., choose between a continuously rising phase characteristic and one that rises and falls, or vice versa. This choice then fixes a and b and thus determines the constant multiplier for the reactance function x. This degree of arbitrariness in the possible resulting phase functions obtainable for a specified attenuation is thereby placed clearly in evidence, and the point in the design procedure at which such a choice must be made is also definitely set forth.

The general character of $(\gamma_1 - \gamma_m)$ as a function of y is illustrated in Fig. 219 for the intervals $\delta = 1$, 2 napiers. In order to recognize a further freedom in design procedure, it should be noted that the rela-

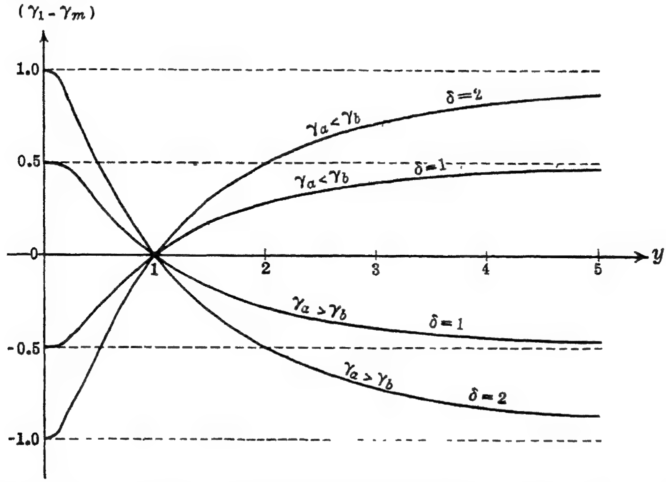


FIG. 219.—Attenuation versus the proportional reactance function (1029) according to the relation (1030) for various attenuation intervals and for the cases of rising and falling characteristics.

tions (1030) and (1030a) remain unchanged if we interchange a and b and invert y with a change in sign. Thus we may interchange a and b in the definitions (1024) provided in the definition (1029) we replace y by $-1/y$.[1] It should also be noted from the curves of this figure that negative geometric symmetry obtains about the point $y = 1$; i.e., equal and opposite ordinates occur for reciprocal values of y, and the

[1] Note that in (1029) this amounts to replacing (x/r) by its negative reciprocal value. That this process leaves the propagation function (both attenuation and phase) unaltered may also be seen directly from (1025) and (1025a). Thus interchanging a and b in (1025) changes the sign of the second term, while replacing the argument of the anti-tangent in (1025a) by its negative reciprocal changes ψ by π radians, which has the effect of again reversing the sign of the second term in (1025) The net effect upon γ, therefore, is nil.

mean value of attenuation occurs for $y = 1$. For a chosen attenuation interval and given attenuation function, the corresponding curve determines the required y as a function of frequency.

The function y may also be directly related to the resulting phase function γ_2. We find from the above

$$\gamma_2 = \tan^{-1}\left(\frac{2 \sinh \frac{\delta}{2}}{y + \frac{1}{y}}\right), \tag{1031}$$

or

$$y = \frac{\sinh \frac{\delta}{2} \pm \sqrt{\sinh^2 \frac{\delta}{2} - \tan^2\gamma_2}}{\tan \gamma_2}, \tag{1031a}$$

in which the $\pm$ signs give reciprocal y-values. The latter relation may be used to determine a necessary y-function for a specified phase function with the corresponding attenuation interval δ as a parameter.[1] Basing the design upon the phase function instead of the attenuation, however, is seldom chosen as a method of approach because a subsequent separate adjustment in the attenuation is not possible.

Assuming that the design has been approached from the attenuation requirement, it is desirable to determine the corresponding phase functions which may be had. This is most readily done by relating attenuation and phase directly by eliminating y, say, between (1030a) and (1031a). This gives

$$\cos \gamma_2 = \frac{\cosh (\gamma_1 - \gamma_m)}{\cosh \frac{\delta}{2}}, \tag{1032}$$

which must be discussed separately for cases I, II, and III.

[1] The reader should note that in cases I and III, a and b are either both positive or both negative so that δ is real. For these cases, therefore, γ_2 is restricted by the condition $\tan\gamma_2 \leqq \sinh \delta/2$. This is to be expected since these are the cases for which the phase function has a maximum or minimum, the latter value being determined from the equal sign in this condition. In case II, on the other hand, a is positive and b negative, so that $\delta = \delta' \pm j\pi$, and (1031$a$) becomes

$$y = \pm j \frac{\cosh \frac{\delta'}{2} \pm \sqrt{\cosh^2 \frac{\delta'}{2} + \tan^2\gamma_2}}{\tan\gamma_2}.$$

Here the phase may increase continuously as we expect from the previous general discussion of this case. Note that in this case y is a pure imaginary function, as is also evident from the attenuation relations (1030) and (1030a). The reactance x, however, according to (1029) remains a real function.

For case I both a and b are positive and hence δ and γ_m are real. Fig. 220 shows a plot according to (1032) of γ_2 versus $(\gamma_1 - \gamma_m)$ for the case $\delta = 1$. Note that the upper curve corresponds to an increasing attenuation function and the lower one to a decreasing function. This is clear from the direction in which the circular impedance locus in Fig. 218 is traversed. The maximum phase occurs where the attenuation

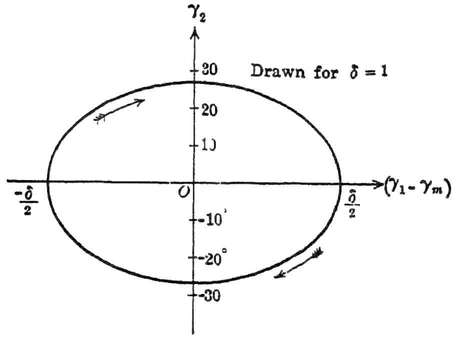


FIG. 220.—Relation between the attenuation and phase functions when the circular impedance locus of Fig. 215 does not enclose the point 1 in the z_a-plane of Fig. 213.

equals the mean value γ_m for the chosen interval. This maximum evidently has the value

$$\gamma_{2\,\max} = \cos^{-1}\left(\frac{1}{\cosh\frac{\delta}{2}}\right), \tag{1032a}$$

and thus depends only upon the attenuation interval. These relations are also indicated in Fig. 218.

For case II, a is positive while b is negative. Hence $\gamma_m = \gamma_m' \pm j\frac{\pi}{2}$, and $\delta = \delta' \pm j\pi$, so that (1032) becomes

$$\cos\gamma_2 = -\frac{\sinh(\gamma_1 - \gamma_m')}{\sinh\frac{\delta'}{2}}. \tag{1032b}$$

This gives rise to the plot of Fig. 220a which is also drawn for $\delta = 1$. The phase increases continuously as the attenuation oscillates between its maximum and minimum values with increasing frequency.

In case III, a and b are both negative. Hence δ is real and $\gamma_m = \gamma_m' \pm j\pi$. The result is the same as pictured in Fig. 220 for case I except for an additive π in the phase.

According to the discussion so far, this approach to the problem of designing a distortion correction network leads one to the collateral problem of finding a non-dissipative two-terminal network for the realization of the reactance function x. Although this problem has been met before (for example, in the determination of susceptance or reactance annulling networks for the Bode method of impedance correction discussed in sections 11 and 12 of Chapter IX), the manner in which it appears here requires several additional comments. When an arbitrary curve for x is given over a specified frequency interval, a physical reactance which approximates this with a prescribed tolerance does not exist unless the curve satisfies certain conditions. One such necessary condition is known to us from our discussion of Foster's and Cauer's reactance theorems in Chapter V. There we saw that a physical reactance must everywhere have a positive slope greater than zero, as was proved in section 3 of Chapter VI. Another necessary condition may readily be shown from the latter discussion. Namely, from equation (535), page 227, we get for a non-dissipative reactance

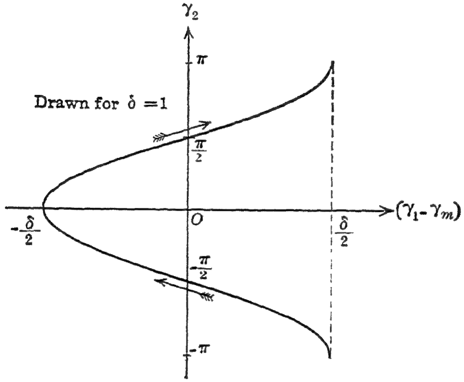


FIG. 220A.—Phase versus attenuation which results when the circular impedance locus encloses the point 1 in Fig. 213.

$$X_{11} = \frac{E_1^2}{4\omega} \cdot \frac{1}{(T_{av} - V_{av})},$$

and combining this with the result (537a), page 229, we find

$$\left(\frac{dX_{11}}{d\omega} - \frac{X_{11}}{\omega}\right) = \frac{2E_1^2V_{av}}{4\omega^2(T_{av} - V_{av})^2},$$

which must always be positive (E_1 is taken as reference and hence is real). Thus we have the result

$$\frac{dX_{11}}{d\omega} > \frac{X_{11}}{\omega} \tag{1033}$$

for a physical reactance. When a curve for X_{11} is specified this means that *the slope of this curve at any point must be greater than that of a*

straight line drawn from the origin to this point. Unless the given curve conforms to this requirement (sufficiently in view of a stated tolerance in the degree of approximation) no physical network can be found. Although the condition of positive slope and that stated by (1033) are necessary, they are very likely not sufficient. A complete solution of this problem is not available at the present time, but it is fairly safe to assume that the above conditions are probably the most essential in the sense that if the given curve satisfies them we may proceed in an attempt to find a network with good chances of success. A method of procedure is first to remove the poles of the function within the prescribed frequency range (if there are any) by means of the method of Foster's reactance theorem.[1] A possible remaining zero in this range may be removed in the same manner, regarding it as a pole of the corresponding susceptance function. The remainder, which is then free from zeros and poles, may be approximated by a network having zeros and poles lying outside the prescribed frequency range, the locations of these serving as parameters by means of which a cut-and-try adjustment may be carried out.

An alternative procedure which is quite effective in simple cases is to set down by inspection the structure for the reactance x which has potential possibilities for meeting the prescribed form of the attenuation function, and demanding agreement at a number of chosen frequencies by writing specific equations involving the functional form of x. The number of these equations, of course, must equal the assumed number of parameters in z_a. The method, for practical reasons, is limited to cases where the required x is simply a coil or condenser, or possibly a resonant or anti-resonant component.

As an illustration let us consider the practical case of a network required to correct for distortion in a 50-mile length of 16-gauge cable with B-22 loading[2] and mid-coil terminations. The nominal value of the characteristic impedance is 800 ohms. Although the reflection losses due to a constant resistance termination are relatively small, they

[1] The parameter determination for an anti-resonant component representing such a pole is made in terms of the residue of the function in that pole according to eq. (468), p. 194, and the residue is determined from the slope of the corresponding susceptance function at this point (which is a zero for the susceptance) according to eq. (465), p. 193. This slope can be measured graphically by plotting the susceptance as well as the required reactance function. The problem, of course, may also be approached from the required susceptance by removing its poles in the specified frequency range as resonant components in parallel. The parameters in the latter are determined from the slope of the reactance in its zeros according to eqs. (471) and (472), p. 194.

[2] 22 milihenry coils spaced at intervals of 3000 ft. Data taken from A. T. & T. Co. notes.

do affect the net result somewhat at the lower and upper limits of the frequency range which is assumed to extend from 30 to 8000 cycles per second. The characteristic necessary to correct for the resultant attenuation over this range[1] is shown by the solid curve in Fig. 221, it being understood that the average level is arbitrary and may finally be placed at any other value.

Except for the peak at the lower end of the range, the characteristic is quite a simple one. A glance at the decreasing curves of Fig. 219 suggests that we attempt to meet the requirements of this problem by

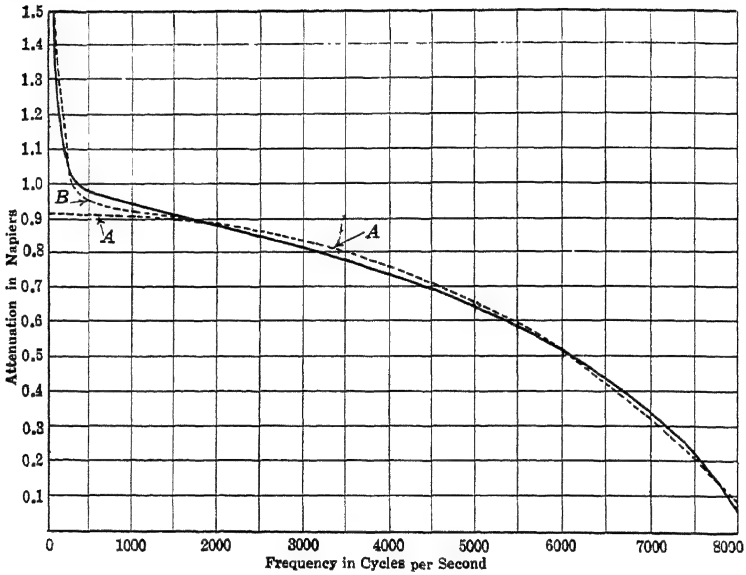


FIG. 221.—Attenuation corrective characteristic for 50 miles of 16 gage cable with B-22 loading terminated at mid-coil in resistances of 800 ohms.

breaking the given characteristic into the sum of two parts, and the corrective network correspondingly into a cascade of two, one of which meets the specified attenuation at the high-frequency end after the fashion of the dotted curve A in Fig. 221, and another which approximates the remainder at the low-frequency end in a manner similar to that shown by the dotted curve B. The requirements of network A then roughly specify a y-function which increases slowly with frequency at the low end and rapidly as the upper end is approached, i.e., somewhat like the variation of the reactance of an anti-resonant component whose pole lies near but to the right of the upper frequency limit. Such

[1] See problem 4–9, p. 182.

a component for the reactance x, therefore, potentially meets the requirements of network A, at least to a first approximation. The form of part B, according to the curves of Fig. 219, may possibly be met by a uniformly increasing y-function, i.e., by an inductance alone for the reactance x.[1]

With this choice of networks for parts A and B, the detailed approximation process may be formulated as follows: If we write for y

$$y = g\nu \tag{1034}$$

where g is a constant multiplier and ν a function proportional to the

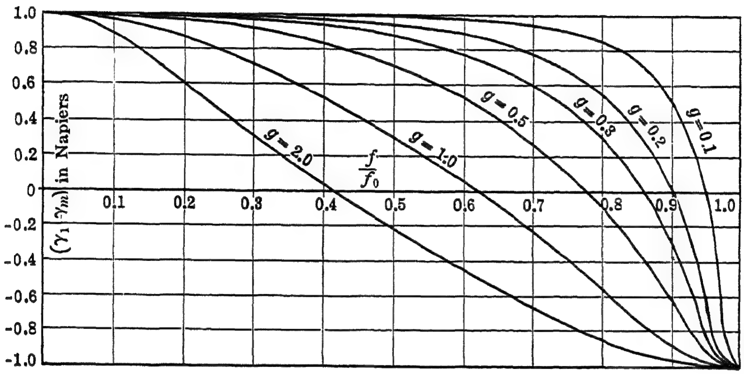


FIG. 222.—Attenuation characteristics of Fig. 219 versus the frequency ratio f/f_0 when the function y is determined according to (1034) and (1036).

reactance x (or proportional to the negative reciprocal of x in case y is later replaced by $-1/y$), and if we let

$$\left.\begin{aligned} h &= e^{\gamma_1} \\ x_1 &= a^2 \\ x_2 &= abg^2 \\ x_3 &= \frac{ag^2}{b} \end{aligned}\right\}, \tag{1034a}$$

then either of the relations (1030) or (1030a) may be written in the form

$$x_1 + \nu^2 x_2 - h^2\nu^2 x_3 = h^2. \tag{1035}$$

[1] The decreasing characteristics of Fig. 219 result when $|a| > |b|$. With the definitions (1024) for a and b, this is possible only in case II for which the locus encloses the point 1. In order to realize this condition for cases I and III, the definitions for a and b must be interchanged and y replaced by its negative reciprocal; i.e., in the present case x will become a resonant component for network A and a condenser for network B.

This equation may be written at three chosen frequencies and the resulting set solved simultaneously for x_1, x_2, and x_3, from which a, b, and g are then obtained.

For the approximation of the characteristic of Fig. 221 at the high-frequency end (network A), we then let

$$\nu = \frac{\dfrac{f}{f_0}}{1 - \left(\dfrac{f}{f_0}\right)^2}, \tag{1036}$$

which has the functional form of an anti-resonant component where f_0 is the anti-resonant frequency and f the variable frequency. Fig. 222

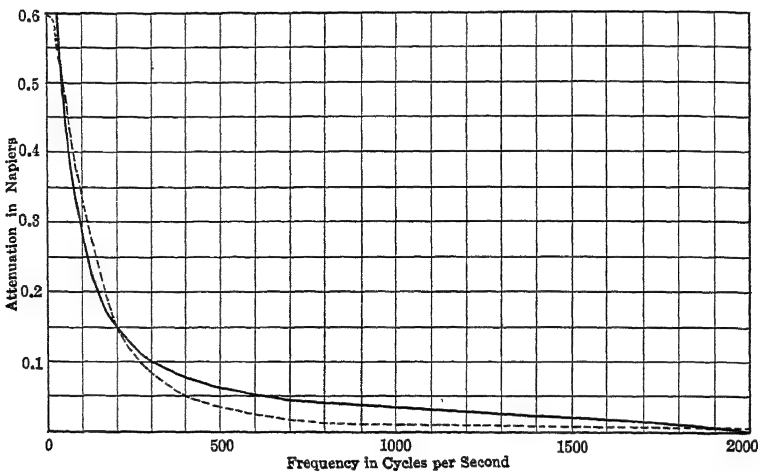


FIG. 223.—Remainder of desired attenuation function in Fig. 221 after the characteristic of network A is subtracted.

shows a family of curves of $(\gamma_1 - \gamma_m)$ versus f/f_0 according to (1030) for various values of g and for an attenuation interval of 2 napiers. These illustrate the general form of results obtainable from such a frequency function, and also enable us to estimate a value for f_0 in our present example. Thus the choice $f_0 = 10{,}000$ seems reasonable. Demanding exact agreement at the frequencies 2000, 6000, and 7800, the resulting equations (1035) when solved simultaneously yield

$$\left.\begin{aligned} a &= 2.17;\ \gamma_a = 0.915 \\ b &= 0.687;\ \gamma_b = -\ 0.235 \\ g &= 0.749 \end{aligned}\right\} \tag{1037}$$

The fact that b comes out less than unity and hence γ_b negative means that we shall have to raise the mean level of our characteristic sufficiently so that no negative values of attenuation result. Since it is desirable to keep the mean level as low as possible (in the interest of maintaining a low resultant average loss), b may be set equal to unity ($\gamma_b = 0$) and a determined so as to keep the same a/b ratio as in (1037).

With the values (1037) we find from (1030) and (1034)

$$e^{\gamma_1} = \sqrt{\frac{4.716 + 0.837\nu^2}{1 + 1.772\nu^2}}$$

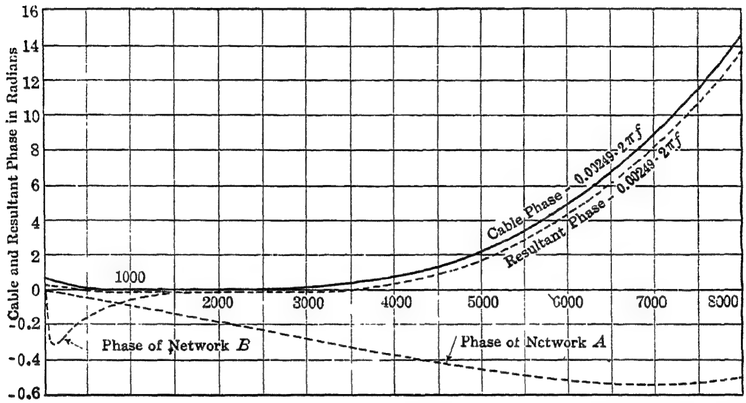


FIG. 224.—Difference between the cable phase and a linear phase function representing a delay of 2.49 miliseconds. Dotted lines are for the attenuation equalizer and for the resultant phase.

This results in the dotted curve A in Fig. 221 which deviates above 2000 cycles from the desired characteristic by not more than 0.026 napier or about 0.23 db, and we may consider it a sufficiently good approximation.

The remainder of the desired attenuation characteristic for the region from zero to 2000 cycles is shown by the solid curve in Fig. 223. For the approximation of this we set $\nu = f$. Then writing (1035) at the frequencies f equals 40, 200 and 2000 cycles, we find from a simultaneous solution

$$\left.\begin{aligned} a &= 1.807;\ \gamma_a = 0.597 \\ b &= 0.998;\ \gamma_b = -0.002 \\ g &= 0.901\cdot 10^{-2} \end{aligned}\right\}. \qquad (1037a)$$

With these values (1030) becomes

$$e^{\gamma_1} = \sqrt{\frac{3.266 + 1.467\cdot 10^{-4}f^2}{1 + 1.472\cdot 10^{-4}f^2}},$$

which gives rise to the dotted curve of Fig. 223 or curve B of Fig. 221. This deviates from the desired characteristic by a maximum of 0.03 napier or 0.26 db, which we shall also consider sufficiently good.

The completion of the equalizer design now depends upon the phase characteristic of the cable. The solid curve of Fig. 224 shows the difference between the cable phase and a straight line representing a delay of 2.49 miliseconds.[1] Although there is a slight deviation from linearity at the low-frequency end, the principal deviation appears at the higher frequencies. Placing the impedance loci of networks A and B so as to

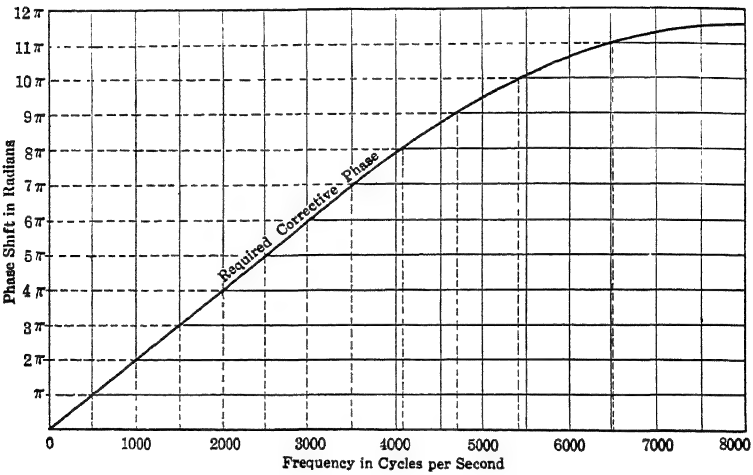


FIG. 225.—Required corrective phase characteristic for the cable circuit to which Fig. 221 applies, after the phase of the equalizer is taken into account.

correspond to cases I or III gives rise, according to (1032), to the dotted characteristics of Fig. 224. Network B is thus seen to produce an improvement at the low-frequency end which may be considered sufficient. Network A produces a slight improvement at the high-frequency end, but still leaves a considerable phase correction to be made there.[2] A phase characteristic which corrects the resultant of Fig. 224, and which has the positive slope required for the physical realization of a corresponding corrective network, is shown in Fig. 225.

[1] This difference is plotted rather than the phase itself in order to make the deviation from linearity stand out more prominently.

[2] It is readily seen that case II does not offer the same potential improvements in the phase characteristic. For this reason, and for the reason that in cases I and III the resultant lattice may be converted to a more economical equivalent bridged-T (as will be shown later), the loci for networks A and B will be placed according to the latter cases.

Since the maximum slope of this corrective phase corresponds to a delay of one milisecond, the net delay of the cable plus its corrective networks will be the sum of this delay and the 2.49 miliseconds already subtracted from the phase characteristic of the cable itself.

For the detailed determination of network A we have by (1037)

$$\frac{a}{b} = 3.16.$$

According to case I both a and b are to be taken positive. Letting

$$a = 3.16;\ b = 1,$$

it follows that

$$\cosh\frac{\gamma_a}{2} = 1.17;\ \cosh\frac{\gamma_b}{2} = 1,$$

so that with (1029), (1034), and (1036) we get

$$y = 0.749\nu = \frac{0.749\dfrac{f}{f_0}}{1 - \left(\dfrac{f}{f_0}\right)^2} = 1.17\left(\frac{x}{r}\right),$$

whence

$$\frac{x}{r} = \frac{0.749}{1.17} \cdot \frac{\dfrac{f}{f_0}}{1 - \left(\dfrac{f}{f_0}\right)^2}.$$

Since (1024) cannot be satisfied this way we may interchange the values of a and b thus

$$b = 3.16;\ a = 1,$$

whence

$$\frac{x}{r} = -\frac{1.17}{0.749} \cdot \frac{1 - \left(\dfrac{f}{f_0}\right)^2}{\dfrac{f}{f_0}}.$$

By (1024) we then have

$$r_0 = 0;\ r = 0.5192,$$

and

$$x = \frac{1 - \left(\dfrac{f}{f_0}\right)^2}{-1.233\left(\dfrac{f}{f_0}\right)}.$$

Recalling that $f_0 = 10{,}000$, the resulting network for z_a becomes that shown in Fig. 226(a) in which the parameter values are indicated in ohms, henries, and farads.

For network B we have by (1037a)

$$\frac{a}{b} = 1.81.$$

If we let

$$a = 1.81;\ b = 1,$$

we get

$$\cosh\frac{\gamma_a}{2} = 1.044;\ \cosh\frac{\gamma_b}{2} = 1,$$

and since in this case $\nu = f$, (1029) and (1034) give

$$y = 0.901\cdot 10^{-2}\cdot f = 1.044\left(\frac{x}{r}\right),$$

whence

$$\frac{x}{r} = 0.863\cdot 10^{-2}\cdot f.$$

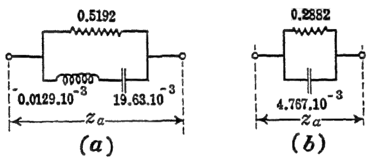


FIG. 226.—Lattice impedances for the equalizer networks A and B.

In order to satisfy (1024), however, we must again interchange a and b, and let

$$b = 1.81;\ a = 1,$$

whence

$$\frac{x}{r} = \frac{-1}{0.863\cdot 10^{-2}\cdot f}.$$

The relations (1024) then yield

$$r_0 = 0;\ r = 0.2882,$$

so that

$$x = \frac{-1}{2.993\cdot 10^{-2}\cdot f}.$$

The resulting network for z_a is shown in part (b) of Fig. 226. Since the corresponding impedances z_b are the reciprocals of these with respect to unity, the final lattices on a one-ohm basis are readily obtained.

It is of interest now to recognize that the lattice resulting from the design of an equalizer may under certain circumstances be converted into an equivalent bridged-T with a consequent saving in the required number of network parameters. In order to investigate this possibility let us consider the equivalence of the networks shown in Fig. 227. By means of the bisection theorem (see p. 439) we have

$$\left.\begin{aligned}\frac{1}{z_a} &= \frac{1}{z_1} + \frac{1}{z_3}\\ z_b &= z_1 + z_2\end{aligned}\right\}. \tag{1038}$$

Since $z_a z_b = 1$, this gives

$$\left.\begin{aligned}z_2 &= \frac{1}{z_a} - z_1\\ \frac{1}{z_3} &= \frac{1}{z_a} - \frac{1}{z_1}\end{aligned}\right\}. \tag{1038a}$$

Thus we see that if z_a has a parallel resistance we may attempt to take a portion of this out as z_1 and its inverse as $1/z_1$, leaving the remainder to represent z_2 and $1/z_3$, respectively. The conditions under which this is possible are readily given. Suppose z_a consists of the resistance r_a' in parallel with a remaining impedance z_a'; i.e., suppose we have

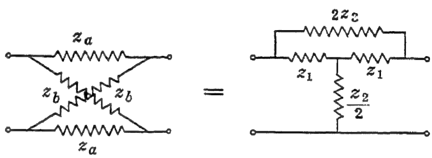


FIG. 227.—Lattice and bridged-T networks whose equivalence relations are expressed by (1038) and (1038a).

$$\frac{1}{z_a} = \frac{1}{z_a'} + \frac{1}{r_a'}; \tag{1039}$$

then if we let

$$z_1 = \rho, \tag{1039a}$$

a positive real quantity, we see that z_2 and z_3 are physical if

$$\frac{1}{r_a'} \geqslant \rho, \text{ and } \frac{1}{r_a'} \geqslant \frac{1}{\rho} \tag{1039b}$$

respectively. The simultaneous fulfilment of these two conditions evidently requires that

$$r_a' \leq 1, \tag{1040}$$

which thus becomes the necessary and sufficient condition for the physical realization of an equivalent bridged-T on this basis.

From our previous discussion of impedance loci it is now interesting to note that the condition (1040) imposed upon the impedance z_a of the form (1039) is equivalent to stipulating that the loci of z_a and z_b shall lie wholly in the regions A or A' respectively, shown as shaded areas in the z_a-plane of Fig. 217 or in the e^γ-plane of Fig. 217a. We conclude, therefore, that *an equivalent bridged-T exists for the cases I or III discussed above, but not for case II.* Thus in this method of design the limitation of the bridged-T over the lattice network lies only in the fact that a continuously increasing phase characteristic cannot be had. *The possibilities of attenuation equalization are the same for the bridged-T as for the lattice.*

In connection with the above argument it should be recognized, of course, that if z_a has a *series* instead of a *parallel* resistance (like r_0 in Fig. 214) then z_b (the reciprocal of z_a) has a parallel resistance, and we need merely interchange these impedances in order to apply the above method directly. The entire argument can be carried out also in terms of an assumed series resistance in the network of z_b. We shall let the reader do this as an exercise.

When (1040) is satisfied, any value of ρ may be chosen within the range (1039b), but the limiting values lead to networks with the least number of resistances. Thus in particular if we set

$$z_1 = r_a', \tag{1041}$$

we get

$$\left.\begin{aligned} z_2 &= \frac{1}{z_a'} + \frac{1 - r_a'^2}{r_a'} \\ \text{and} \qquad z_3 &= z_a'. \end{aligned}\right\}, \tag{1041a}$$

while if we put

$$z_1 = \frac{1}{r_a'}, \tag{1042}$$

we have

$$\left.\begin{aligned} z_2 &= \frac{1}{z_a'} \\ \text{and} \qquad \frac{1}{z_3} &= \frac{1}{z_a'} + \frac{1 - r_a'^2}{r_a'} \end{aligned}\right\}. \tag{1042a}$$

The saving in elements over the lattice form is quite apparent.

Using the results (1041) and (1041a) in connection with the impedances of Fig. 226 for networks A and B of the above design, the resulting equalizer on a one-ohm basis takes the form shown in Fig. 228. The parameter values are given in ohms, henries, and farads. Con-

version to the 800-ohm basis merely requires multiplying all resistance and inductance values by 800 and all capacitance values by 1/800.

This illustrative example shows how the above reasoning in terms of impedance loci and the chart of Fig. 213 enables one to lay out the approximate network requirements of a particular problem quite readily. In cases where the tolerances are lenient a design may be roughed through fairly rapidly. Thus, for example, where simply a continuously decreasing or increasing attenuation characteristic is required over the audio-frequency range, a resonant or anti-resonant component for x evidently meets the situation, and the curves of Fig. 222 show in more detail the variety of results obtainable. When the required curve has several maxima and minima, a continuously varying portion can first be removed and the remainder approximated by a second network with a reactance x in the form of several anti-resonant components in series or resonant components in parallel with poles and zeros within the prescribed frequency range, and the resistances r_0 and r determined from the desired extremal values.

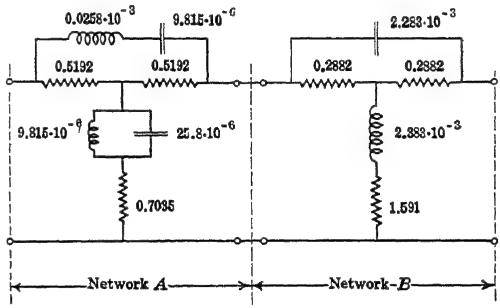


FIG. 228.—Complete attenuation equalizer after the lattices for networks A and B are converted to equivalent bridge-T forms.

Since the extremal values within which the attenuation function varies depend in a simple manner upon the resistances r_0 and r alone, a network may readily be designed for which these values are adjustable by introducing suitably variable resistances. In this respect we note from (1024) that if r_0 and r are changed to the values r_0' and r' such that

$$r_0' = \frac{1}{r_0 + r}; \; r_0' + r' = \frac{1}{r_0},$$

the values of a and b become

$$a' = -b; \; b' = -a.$$

In case II this change may be made continuous, with the result that a continuously decreasing attenuation function (for example) passes through the constant form to that of a continuously increasing function

with the same extremal values. It is easy to see how a variety of results may be reached by modifications of this process. Such networks may be useful as means for "tone control" in sound reproduction or recording equipment.

In this connection it is also interesting to note from (1030) that replacing y by $-1/y$ replaces the function $(\gamma_1 - \gamma_m)$ by its negative. For fixed values of a and b this is accomplished, according to (1029), by reciprocating the ratio (x/r) so as to obtain $(x/r)'$ fulfilling the relation

$$\left(\frac{x}{r}\right) \cdot \left(\frac{x}{r}\right)' = -\left(\frac{\cosh \frac{\gamma_b}{2}}{\cosh \frac{\gamma_a}{2}}\right)^2. \tag{1043}$$

Reference to Fig. 220 shows that in cases I or III (to which this figure applies) the phase function is replaced by its negative. Thus in cases I or III the process of reciprocation indicated by (1043) leads to the complementary network; i.e., the two networks in cascade form a distortionless system. In case II the process also leads to the complementary network except for the fact that a linear component in the phase characteristic still remains. In this case the two networks in cascade again form a distortionless system, but one with a finite instead of a zero delay.

We return now to a more detailed consideration of the phase characteristic and a discussion of means for counteracting the distortion caused by it. As pointed out above, the phase function necessary to correct for this kind of distortion in the cable problem is shown in Fig. 225. We shall use this example as a means for illustrating the general method of attack. Although the latter has much in common with that used in the design of artificial lines or linear phase-shift networks (Chapter VI, sections 1 and 2) as well as with the process of meeting the phase requirements of filters by the Bode method (Chapter X, section 9), we shall repeat the essential features in some detail here.

The all-pass, constant-impedance network used for this purpose is a special case of that treated above for attenuation equalization when z_a and z_b are assumed to be pure reactances. In this case $z_a = jx_a$, and (1019) reads

$$\left.\begin{aligned} x_a &= \tan \frac{\gamma_2}{2} \\ \gamma_2 &= 2 \tan^{-1} x_a \end{aligned}\right\}, \tag{1044}$$

where γ_2 is the desired phase function. Since the tangent function has uniformly spaced zeros and poles at integer multiple increments of $\pi/2$

in its argument, the initial procedure is quite clear. With γ_2 specified over a given frequency range, the locations of the zeros and poles of the reactance x_a within this range are immediately given. Referring to Fig. 225, the γ_2-scale is divided into increments of π radians, and by projecting these through the desired phase curve the corresponding critical frequencies of x_a are located. Recognizing that the origin is a zero and that zeros and poles alternate, we have as a preliminary requirement for x_a the functional form

$$x_a = \frac{h \cdot f \cdot \left(1 - \frac{f^2}{f_3^2}\right)\left(1 - \frac{f^2}{f_5^2}\right) \cdots \left(1 - \frac{f^2}{f_{11}^2}\right)}{\left(1 - \frac{f^2}{f_2^2}\right)\left(1 - \frac{f^2}{f_4^2}\right) \cdots \left(1 - \frac{f^2}{f_{12}^2}\right)}, \tag{1045}$$

where h is a constant multiplier, f stands for frequency, and f_2, f_3, f_4, . . . are the critical frequencies graphically found to be

Zeros	Poles
$f_3 = 1000$	$f_2 = 500$
$f_5 = 2000$	$f_4 = 1500$
$f_7 = 3000$	$f_6 = 2500$
$f_9 = 4080$	$f_8 = 3500$
$f_{11} = 5440$	$f_{10} = 4680$
	$f_{12} = 6520$

(1045a)

The constant multiplier h determines the behavior of x_a for small values of f and hence must be fixed, according to (1044), by the nature of the required phase function in the vicinity of the origin. If we wish here to demand exact agreement between the desired phase and that of the corrective network, (1044) is considered for small values of x_a and γ_2, the latter in this example being given by $\gamma_2 = 2\pi \cdot 10^{-3} f$. Thus we have

$$h \cdot f = \pi \cdot 10^{-3} \cdot f,$$

so that

$$h = \pi \cdot 10^{-3}. \tag{1045b}$$

The slope of the γ_2-curve at the origin thus determines h.

In order to examine the degree of approximation afforded by the x_a-function (1045) with the parameter values (1045a) and (1045b), it is effective (as in the discussion of artificial line design in Chapter VII, pp. 260-271) to study the ratio $\left(x_a / \tan \frac{\gamma_2}{2}\right)$ over the essential frequency range. Values of this ratio may be calculated at a number of frequencies between zeros and poles. The ratio being a continuous

function of frequency, a relatively small number of such values suffice to draw a smooth curve such as the upper dotted one in Fig. 229 which pertains to our present example. For perfect agreement between desired and resulting characteristics, this ratio should, of course, equal unity over the entire range. The above determination of h makes the ratio unity at the origin, but a considerable deviation becomes evident toward the high-frequency end. The effect which such a deviation in this ratio has upon the resulting phase characteristic of the network has been discussed in section 2 of Chapter VII on pages 270-271, and also in section 9 of Chapter X on pages 416-417. Thus if the deviation

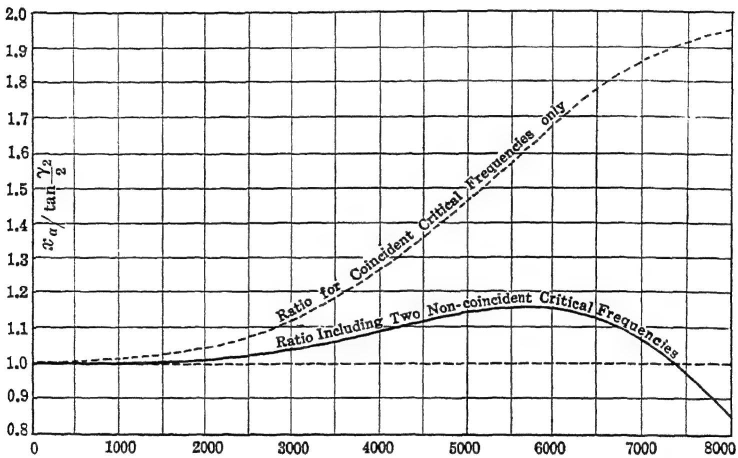


FIG. 229.—Ratio of the actual lattice reactance x_a to its required functional form $\tan \gamma_2/2$ for the preliminary form (1045) (dotted curve) and for this form augmented by the factor (1045*c*).

is denoted as a decimal error ϵ we have for the corresponding deviation in the phase function

$$d\gamma_2 = \epsilon \sin \gamma_{20} \tag{1046}$$

where γ_{20} may be considered to be the required (undistorted) phase function. Over small regions where the latter is substantially linear, the deviation in the resulting phase function from the desired one is a sinusoid with maximum amplitude equal to the decimal deviation in the ratio $\left(x_a/\tan \frac{\gamma_2}{2}\right)$ from unity. Although the derivation of (1046) assumes small deviations, this result may still be applied with surprising accuracy in the case of deviations ϵ as large as 100 per cent. The phase function of the corrective network is thus seen to oscillate about the

desired curve (a half-period corresponding to each increment of π in γ_2) with an amplitude (in radians) equal to the error ϵ as indicated by the ratio curve plotted for our present example in Fig. 229.

The preliminary x_a-function (1045), therefore, leads to a network whose phase characteristic oscillates about the required curve with an increasing amplitude, the latter reaching 0.94 radian (0.299π radian or 53.9°) at the upper essential frequency. As pointed out in Chapter XI, phase distortion (delay distortion) is due chiefly to variations in the *slope* of the phase characteristic, and in section 2 of Chapter VII (p. 271) the decimal variation in the slope for the type of network discussed here, evaluated in terms of ϵ and denoted as the index of quality, was found to be given by

$$\delta_{\gamma_2} = \epsilon \cos \gamma_2. \tag{1046a}$$

The significant feature of this result is the fact that the maximum value of the decimal variation in slope equals ϵ as indicated by the ratio $\left(x_a/\tan \frac{\gamma_2}{2}\right)$. In our present example this reaches the value of 0.942 at 8000 cycles per second based on a norm (average slope of the phase characteristic) of 10^{-3} second. This means that the design so far leads to a maximum variation in delay of 0.942 milisecond (the basic delay of the 50 miles of loaded cable plus its corrective networks, it will be recalled (see p. 533), is 3.49 miliseconds).

This rather considerable variation in delay we shall consider too great, and our preliminary design, therefore, insufficient. Recalling the process used in the design of artificial lines on the lattice basis, the method of improving this situation is quite clear. Namely, we must add more zeros and poles (lying outside the essential frequency range) to the reactance function (1045), and utilize the additional critical frequencies thus introduced as design parameters in producing a better agreement between the ratio $\left(x_a/\tan \frac{\gamma_2}{2}\right)$ and unity. Each additional critical frequency introduces a factor of the form $\left(1 - \frac{f^2}{f_\nu^2}\right)$ alternately in the numerator and denominator of the expression (1045). Suppose we introduce one more zero and one more pole. Then x_a, and hence the ratio $\left(x_a/\tan \frac{\gamma_2}{2}\right)$, becomes multiplied by the factor

$$\frac{1 - \dfrac{f^2}{f_{13}^2}}{1 - \dfrac{f^2}{f_{14}^2}}. \tag{1045c}$$

The function given by the dotted curve of Fig. 229 multiplied by this factor then equals the new ratio $\left(x_a/\tan\frac{\gamma_2}{2}\right)$. By means of a few trials we find that by setting

$$\left.\begin{aligned} f_{13} &= 10{,}500 \\ f_{14} &= 40{,}000 \end{aligned}\right\}, \tag{1045d}$$

the resultant ratio is given by the solid curve in Fig. 229 for which $\epsilon = 0.15$, and hence the maximum variation in delay is 0.15 milisecond (since the norm is given by the maximum slope of the characteristic of Fig. 225 and hence is one milisecond in this case). If the corresponding phase function were plotted in Fig. 225, it would show an oscillatory variation with a maximum deviation from the desired curve of 0.15 or 0.048π radian which is so small as not to be noticeable with the scale used in this figure. Although any desired further improvement may be had by continuing this process, we shall not carry it any farther here. The network for z_a thus consists of seven anti-resonant components in series or any of its potential equivalents as given by the Foster and Cauer methods discussed in Chapter V. The parameter values in these networks are easily found by the methods presented there, and, therefore, we may consider our present problem solved.

In the form of a lattice this network involves 56 elements, half of which are coils. The chief reason for the large number of elements is due to the large required maximum phase shift of almost 12π radians. This is inherent in the type of problem considered here, namely, that of the loaded cable which is a system having a high normal delay. An unloaded line or cable would present an entirely different picture in this respect.

Several possibilities suggest themselves toward reducing the number of required network elements. We may make use of the lattice equivalents discussed with regard to the lattice-type filter in section 10 of Chapter X (pp. 433–434). Thus making use of the equivalent shown in Fig. 179, p. 433, cuts the number of elements almost in half, although it does require close-coupled coils which can be realized only approximately. In this respect the equivalent network illustrated in Fig. 181 embodying the transformer equivalent of Fig. 180 is more promising since it involves only one mutual inductance. Moreover, the value of the coupling coefficient is controllable and thus easily realizable. This equivalent network is further interesting from the standpoint of the relation it bears to the ordinary bridged-T. For a more detailed discussion of this point we may make use of the network of the above example. Potential networks for z_a and z_b in this case are illustrated in

Fig. 230. Since these reactances are inverse, the inductance values in z_a become capacitance values in z_b, and vice versa. Following the ideas involved in Figs. 180 and 181 leads to the equivalent network shown in Fig. 230a for which we must have

$$\left.\begin{array}{l} L + M = pc_{15};\ p \leq 1 \\ L - M = ql_1;\ q \geq 1 \end{array}\right\}, \tag{1047}$$

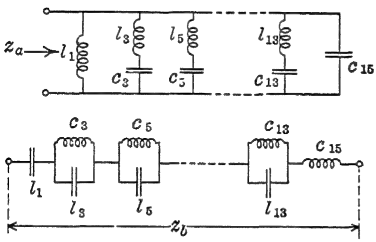


FIG. 230.—Potential networks for the lattice reactances of the phase corrective network specified by the curve of Fig. 225.

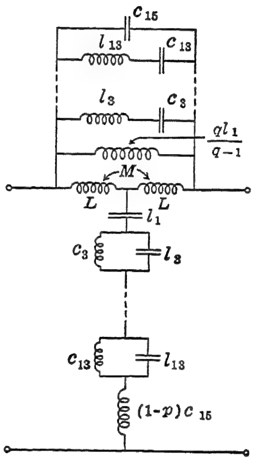


FIG. 230A.—Equivalent network for the lattice whose reactances are shown in Fig. 230.

where p and q are numerics whose values are restricted as indicated. It follows that

$$\left.\begin{array}{l} L = \dfrac{1}{2}(pc_{15} + ql_1) \\[2ex] M = \dfrac{1}{2}(pc_{15} - ql_1) \end{array}\right\}. \tag{1047a}$$

The values of L and M are thus adjustable except for the restrictions on p and q. The resulting network of Fig. 230a is a bridged-T which differs from the ordinary only in the appearance of the mutual inductance M. Here it is logical to attempt making M zero. The condition for this is evidently that

$$pc_{15} - ql_1 = 0. \tag{1048}$$

The restrictions on p and q however, make this possible only when

$$\frac{l_1}{c_{15}} \leq 1, \tag{1048a}$$

which thus becomes the condition under which the lattice in this case is reducible to an ordinary bridged-T.

The situation presented by this example suggests that we investigate the possibility of finding an ordinary bridged-T for the lattice from a somewhat more general point of view. For this purpose we return to the network equivalence illustrated in Fig. 227 and expressed analytically by the relations (1038a). Here we note that if the reactance z_a of the lattice can be considered as the parallel combination indicated by

$$\frac{1}{z_a} = \frac{z_1}{p} + \frac{q}{z_1} + \frac{1}{z_a'}, \tag{1049}$$

i.e., as a parallel combination of a reactance and its proportionate inverse, plus a remainder z_a', then the reactances z_2 and z_3 of the bridged-T become

$$\left.\begin{aligned} z_2 &= \frac{1-p}{p} z_1 + \frac{q}{z_1} + \frac{1}{z_a'} \\ \frac{1}{z_3} &= \frac{z_1}{p} + \frac{q-1}{z_1} + \frac{1}{z_a'} \end{aligned}\right\}. \tag{1049a}$$

In order that the latter be physical it is necessary that $p \leq 1$ and $q \geq 1$. The conditions under which z_a may be broken into parallel components according to (1049) may be generally formulated by recalling the detailed discussion of Foster's reactance theorem. In order for it to be possible to subtract from $1/z_a$ part of a reactance z_1 and its proportionate inverse and leave a physical remainder, we find that:

(a) *z_1 may be critical only at the zeros of z_a.*
(b) *z_a must have zeros at the origin and at infinity.*
(c) *At the zeros of the functions z_1 and $1/z_1$ their slope must be less than (or equal to) that of z_a.*

Requirement (a) is necessary in order that poles of both z_1 and $1/z_1$ be contained in z_a, and (b) is an obvious consequence of this requirement. The condition (c) must be fulfilled in order that the residues of $1/z_1$ and of z_1, respectively, be less than those of $1/z_a$ where these functions have coincident poles. This is equivalent to the conditions $q \geq 1$ and $p \leq 1$, respectively, in (1049), which are necessary for the realization of z_2 and z_3 according to (1049a).

In the general case where z_a is given by a structure of n meshes of the form shown in Fig. 230, its functional expression reads

$$z_a = \frac{jh \cdot f \cdot \left(1 - \frac{f^2}{f_3^2}\right)\left(1 - \frac{f^2}{f_5^2}\right) \cdots \left(1 - \frac{f^2}{f^2_{2n-3}}\right)}{\left(1 - \frac{f^2}{f_2^2}\right)\left(1 - \frac{f^2}{f_4^2}\right) \cdots \left(1 - \frac{f^2}{f^2_{2n-2}}\right)} \tag{1050}$$

If we let

$$\frac{z_1}{q} = j\omega l_1, \tag{1050a}$$

then the shunt condenser, which is denoted by c_{2n-1}, becomes that portion from which the inverse reactance may be subtracted, i.e.

$$\frac{1}{jc_{2n-1}\omega} = \frac{p}{z_1}. \tag{1050b}$$

We then have

$$\frac{l_1}{c_{2n-1}} = \frac{p}{q} \leq 1, \tag{1051}$$

as the necessary and sufficient condition for the realization of the equivalent bridged-T in this case. This is identical with the condition (1048a) found previously from an entirely different line of attack. By means of (1050) this condition may now be put into an alternative form which is somewhat more enlightening. Namely, since the inductance l_1 and the capacitance c_{2n-1} determine the function z_a for very small and very large frequencies, respectively, we see that

$$\left.\begin{aligned} l_1 &= \frac{h}{2\pi} \\ \frac{1}{c_{2n-1}} &= 2\pi h \cdot \left(\frac{f_2 \cdot f_4 \cdot \cdots \cdot f_{2n-2}}{f_3 \cdot f_5 \cdot \cdots \cdot f_{2n-3}}\right)^2 \end{aligned}\right\}, \tag{1052}$$

and hence the condition (1051) is equivalent to

$$h \leq \frac{f_3 \cdot f_5 \cdot \cdots \cdot f_{2n-3}}{f_2 \cdot f_4 \cdot \cdots \cdot f_{2n-2}}. \tag{1051a}$$

In order to see how this restricts the type of phase function obtainable, we note that for very small frequencies we have with the use of (1044)

$$h = \pi t_{d0} \tag{1053}$$

where t_{d0} is the slope of the phase function (versus $\omega = 2\pi f$) at the origin. If this function has a constant average slope t_d, then $2\pi f_\nu t_d = \pi(\nu - 1)$, so that

$$f_\nu = \frac{\nu - 1}{2t_d},$$

and we get by combining (1051a) and (1053)

$$\frac{t_{d0}}{t_d} \leq \frac{2}{\pi} \cdot \frac{2 \cdot 4 \cdot \cdots \cdot (2n-4)}{1 \cdot 3 \cdot \cdots \cdot (2n-3)}. \tag{1053a}$$

For the simplest case $n = 2$ this is

$$\frac{t_{d0}}{t_d} \leq \frac{2}{\pi};$$

for $n = 3$

$$\frac{t_{d0}}{t_d} \leq \frac{2}{\pi} \cdot \frac{2}{1 \cdot 3},$$

and so forth. We note that the initial slope of the phase function is less than the average, and that this situation becomes more accentuated as n increases. The resulting phase function in the vicinity of the origin is evidently undulatory in character. A linear phase-shift network, therefore, cannot be realized with the ordinary bridged-T, but it can be had by admitting a mutual inductance in this network as shown in Fig. 230a.[1]

In the discussion of this and similar equivalent lattice networks in the design of filters it was pointed out that practical difficulty is encountered in attempting to build the filter in this form on account of the small tolerances allowable for the elements, particularly when high values of attenuation are involved. In the all-pass structure discussed here, this feature, although less aggravated, may still be present to a certain extent. Therefore, it is again desirable to attempt a decomposition of the resultant complicated lattice or its equivalent into a cascade of simpler structures.

[1] In this connection it is interesting to note that for a linear phase-shift network ($t_{d0} = t_d$) we have from (1052)

$$\frac{l_1}{c_{2n-1}} = \left(\frac{\pi}{2} \cdot \frac{1 \cdot 3 \cdot \cdots \cdot (2n-3)}{2 \cdot 4 \cdot \cdots \cdot (2n-4)}\right)^2.$$

According to (1047a) the necessary coupling coefficient for the mutual inductance is least for $p = q = 1$, and is given by

$$k = \frac{\dfrac{l_1}{c_{2n-1}} - 1}{\dfrac{l_1}{c_{2n-1}} + 1}.$$

For $n = 2, 3, 4, \cdots$, this gives $k = 0.423, 0.695, 0.793, \cdots$. For large values of n this approaches

$$k \cong 1 - \frac{4}{(2n-3)\pi}.$$

Hence the required coupling coefficient becomes larger with an increasing maximum phase shift.

This may be done on the same basis as that used for filter networks. Recalling the discussion of this process in section 10 of Chapter X, we note that the points of infinite attenuation are used as a basis for such a separation into component networks. In the all-pass structure, the frequencies corresponding to infinite attenuation, of course, are all imaginary or complex, but the analytic method still applies.[1] Since

$$\gamma = \ln\left(\frac{z_a + 1}{z_a - 1}\right), \tag{1054}$$

the frequencies of infinite attenuation are the roots of the equation

$$z_a = 1. \tag{1054a}$$

In order to illustrate some of the detailed aspects of this process, suppose we consider a reactance z_a consisting of an inductance l in series with a capacitance c. The characteristic equation (1054a) then reads

$$\frac{1 + lc\lambda^2}{c\lambda} = 1, \tag{1055}$$

where λ has been written for $j\omega$. The roots of this quadratic equation are given by

$$\lambda_{1,2} = \frac{1}{2l}\left(1 \pm \sqrt{1 - \frac{4l}{c}}\right) = \frac{1}{\frac{c}{2}\left(1 \mp \sqrt{1 - \frac{4l}{c}}\right)}. \tag{1055a}$$

When

$$\frac{4l}{c} \leq 1, \tag{1055b}$$

the roots are real. In that case the given lattice may be replaced by a cascade of two with individual reactances given by either

$$z_{a1} = \frac{\lambda}{\lambda_1}; \; z_{a2} = \frac{\lambda}{\lambda_2}, \tag{1056}$$

or

$$z_{a1} = \frac{\lambda_1}{\lambda}; \; z_{a2} = \frac{\lambda_2}{\lambda}, \tag{1056a}$$

[1] It is significant to note that this method of the synthesis of component networks on the basis of their points of infinite attenuation is applicable only to reactive (non-dissipative) structures which are uniquely characterized by two functions such, for example, as the characteristic impedance and attenuation functions. In dissipative structures (such as the equalizer networks) which are characterized over the same frequency range by attenuation, phase, and impedance, this process of decomposition no longer applies.

i.e., by either inductances $l_1 = 1/\lambda_1$; $l_2 = 1/\lambda_2$, or capacitances $c_1 = 1/\lambda_1$; $c_2 = 1/\lambda_2$. These individual lattices are evidently the simplest components to which a more complicated one may be reduced.

When the condition (1055*b*) is not fulfilled, the roots (1055*a*) are in the form of a pair of conjugate complex values. In that case the lattice of the present example cannot be further reduced. Note, however, that a pair of conjugate complex λ-roots can always be realized by a lattice with a reactance z_a consisting of a coil and condenser in series. Namely, if the roots are

$$\lambda_{1,2} = a \pm jb, \tag{1057}$$

then

$$l = \frac{a^2 + b^2}{2a}; \; c = 2a. \tag{1057a}$$

The procedure in the general case is thus clear. Equation (1054*a*), which is preferably written with $j\omega$ replaced by λ, is solved for its λ-roots. The real roots then correspond to the simplest lattice types, while each pair of conjugate complex roots defines a lattice with l and c according to (1057*a*).[1] Components of this type may be replaced by equivalent bridged-T networks following the general method given above.

In the case of complicated $z_a(\lambda)$-functions, the solution of the equation (1054*a*) is a long and tedious process. The possibility exists, of course, of attempting to synthesize the desired phase function directly by combining those of simpler networks on a cut-and-try basis. This in general is also a long and tedious process, so that a certain amount of disagreeable calculation seems to be unavoidable.

3. Compensating and simulating networks. The problem of designing line balances or simulating networks in practice subdivides into two cases, the treatments of which ordinarily differ considerably in their manner of approach. These cases deal respectively with the smooth and the lump-loaded lines.

The treatment for the smooth line follows the logical and direct approach of finding a two-terminal network whose driving-point impedance approximates the characteristic impedance of the line. The behavior of the latter versus frequency was discussed in Chapter III assuming constant line parameters R, L, G, and C. As already pointed

[1] In order that l and c be positive the λ-roots must evidently have positive real parts. On the basis of the separation property of the zeros and poles of the reactance z_a it is possible to show that the equation $z_a(\lambda) = 1$ can have roots with positive real parts only, so that the process outlined here must always lead to physical component networks.

out, the parameters R and G (particularly the latter) are in practice found to vary with frequency, so that the actual dependence of Z_0 upon frequency is modified to some extent. In any case it is a simple matter, however, to determine curves of the magnitude and angle (or real and imaginary part) of Z_0 versus frequency over the range pertaining to the particular application under consideration. Although a complete analytic solution to the problem of finding a two-terminal network with an impedance function whose real and imaginary parts approximate graphical specifications over a limited frequency range is not available, certain empirical cut-and-try processes for accomplishing this in a satisfactory manner have been developed. The most important of these, which is due to Hoyt,[1] will be discussed in outline form only since the presentation of details in such a method necessarily involves more space than can be given to this one item here.

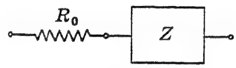


FIG. 231.—Schematic representation of a line balance in terms of a basic resistance element R_o and an excess simulator Z.

Briefly, the complete simulator is initially thought of as being composed of two series parts as illustrated in Fig. 231. Here R_0, called the **basic resistance element,** whose value is taken equal to the nominal characteristic impedance $\sqrt{L/C}$ (non-dissipative value),[2] represents the first approximation to the solution. The remaining discrepancy is represented by the two-terminal impedance Z, which is referred to as the **excess simulator.** In the practical case where $R/L > G/C$, this two-terminal dissipative network must evidently be primarily a capacitative reactance. A general form is illustrated in Fig. 232.

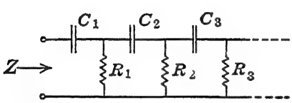


FIG. 232.—Potential network for the excess simulator of Fig. 231 for the uniform line in which the parameters are such that $R/L > G/C$.

Networks of this type, and three other, equivalent fundamental forms were discussed in section 9 of Chapter V, the general functional expression for Z in the case of n meshes being given by equation (508), page 211. A general procedure is to assume a definite number of meshes (this depends upon the degree of precision required), write down the expression (508), and separate into real and imaginary parts (the coefficients in these parts become

[1] "Impedance of Smooth Lines, and Design of Simulating Networks," by Ray S. Hoyt, B.S.T.J., Vol. 2, No. 2, April, 1923, pp. 1–40.

[2] Since this is the minimum value of the real part of Z_0 for the practical case $R/L > G/C$, the remainder, which is contained in Z, becomes positive as it must be if a physical network is to be found for Z. If R/L were less than G/C, then R_0 would be taken equal to $\sqrt{R/G}$, which would then be the minimum value.

functions of the α's and β's in (508)). By a cut-and-try process the parameters in the functions for the real and imaginary parts are determined so that a sufficient agreement with the stipulated curves is obtained. Thus the coefficients in (508) are determined and hence the network also.

In cases not requiring a high degree of precision, this process may be abbreviated by setting up the expressions for the real and imaginary parts of Z in terms of the resistance and capacitance parameters directly. Hoyt works out specific parameter determinations for various practical cases and also gives several equivalent forms for the complete simulative networks.

The treatment for the lump-loaded line follows a somewhat different procedure. As already pointed out, the lump-loaded line is very similar in behavior to the low-pass, constant-k filter. This is particularly true if the amount of loading inductance is large compared with the line inductance, as it is in cables. The procedure for the design of line balances here follows in general the method discussed in section 12 of Chapter IX relative to the termination of filters by the process of impedance correction. A first approximation, it will be recalled, is obtained either by using a fractional series inductance followed by a reactance-annulling or -compensating network and then terminating in a constant resistance equal to the nominal value, or by using a fractional shunt capacitance followed by a susceptance-annulling network and again terminating in the nominal resistance.[1] A higher degree of precision is obtained by the use of additional building-out sections and more elaborate reactance- or susceptance-annulling networks. The details of this process with particular reference to line termination are discussed by Hoyt[2] in a companion paper to the one referred to above

The essential difference between this process and the one discussed for the smooth line lies in the fact that the impedance corrective network is a **transducer** having a band of free transmission over the transmitting range of the line with which it is associated. The final terminating resistance may represent any load matched at the output terminals of the transducer, while the resistance R_0 in the simulator of Fig. 231 cannot in general be considered in this manner.[3]

[1] These fractions are obtained by proper adjustment of the size and spacing in the last loading section.

[2] "Impedance of Loaded Lines, and Design of Simulating and Compensating Networks," by Ray S. Hoyt, B.S.T.J., July, 1924, pp. 414–467.

[3] It is interesting to note, however, that, if R_0 is considered as a load, the network not only simulates the line but also acts approximately as an attenuation equalizer since the impedance Z is a decreasing function with frequency, thus favoring the higher frequencies which are attenuated more strongly by the line.

In the non-dissipative line, the impedance corrective network may be used to join successfully a loaded line to a smooth line. In the dissipative case this is still true so long as the system as a whole is uniformly dissipative, i.e., if all R/L ratios are equal and all G/C ratios are equal for both lines as well as for the coils and condensers in the network joining the two lines. Then the joining network may be looked upon as transforming the lump-loaded line impedance to that of the smooth line, or vice versa. When the two lines do not have the same dissipation ratios, discrepancies arise which must be dealt with by more elaborate methods.

Other methods of designing simulative or corrective networks, such as those discussed by Bartlett,[1] are impracticable chiefly because of their uneconomical use of circuit elements. This is due to the fact that these methods of synthesis depend upon infinite expansions of the functions characterizing line behavior. Although the resulting networks may be broken off after the representation of a finite number of terms, the rate of convergence is determined by the properties of the expansions and is not as rapid as may be had from a method which follows no set rule except the one of approximating the desired characteristics. Furthermore, these methods must of necessity assume the constancy of the line parameters and hence can be used only as first approximations in practical cases where parameter variations are to be considered. As such they may prove quite useful, however, and the student will do well to familiarize himself with this material. The analytic aspects of these methods have a decided similarity to the manipulations involved in carrying out physical realizations according to Foster's and Cauer's theorems regarding driving-point impedances, with which the student is already familiar. A repetition of this material here, therefore, seems superfluous.

[1] A. C. Bartlett, "The Theory of Electrical Artificial Lines and Filters," John Wiley & Sons, 1930, Chapter V, pp. 75–92.

PROBLEMS TO CHAPTER XII

12–1. Construct a system of constant attenuation and constant phase loci in the z_a-plane according to the method indicated in Fig. 213a.

12–2. By means of the chart of the preceding problem and the method of constructing impedance loci, show that the network of Fig. 214 for z_a with simply a coil or a condenser for the reactance x gives rise to attenuation and phase characteristics which tend to correct for the distortion in a long uniform line or cable circuit provided the latter is sufficiently well terminated. Show further that both attenuation and phase correction can be had provided the impedance locus lies wholly to the right or left of the singular point in the z_a-plane. Sketch the lattice structure and its bridged-T equivalents.

12–3. From the results of Problem 3–1 for the open-line data specified there, determine and plot the characteristics which correct the attenuation and phase characteristics for a 100-mile length. Using the network of the preceding problem, determine the parameter values which lead to the best average correction, and determine an equivalent bridged-T network.

12–4. By considering a more elaborate network, obtain corrective characteristics for the line of Problem 12–3 to within a maximum tolerance of 0.5 db over the frequency range 0–4000 cycles per second.

12–5. Repeat Problem 12–4, taking into account the losses and angle due to mismatch at the terminals, assuming the line to be terminated in pure resistances of 700 ohms at either end. Make use of the results of Problem 3–3.

12–6. Repeat Problem 12–5 for a 50-mile length of the cable circuit treated in Problems 3–2 and 3–4. Assume terminal resistances of 500 ohms.

12–7. A cable circuit has the following uniformly distributed parameters

$$R = 85\ 2;\ L = 0.001;\ C = 0.066 \cdot 10^{-6};\ G = 1.585 \cdot 10^{-6}.$$

The circuit is loaded by inserting coils having a resistance of 2.4 ohms and an inductance of 0.044 henry at intervals of 6000 ft. Assuming the parameters to be constant, calculate the insertion loss and angle for a 50-mile length terminated in 800 ohms at either end according to the method outlined in Problem 4–9, and from these data plot the attenuation and phase corrective characteristics over the range 30–5000 cycles per second. Following the method illustrated by the example in the text, carry through a complete corrective network design with a maximum tolerance of 0.5 db in the attenuation and a maximum deviation from constancy in the slope of the phase characteristic of 5 per cent.

12–8. Consider two symmetrical lattice structures with the component impedances z_{a1}, z_{b1} and z_{a2}, z_{b2}, respectively, for which $z_{a1}\, z_{b1} = z_{a2}\, z_{b2} = 1$. If these are connected in cascade, show that, since the overall propagation function must equal the sum of the individual ones, the impedances of the single symmetrical lattice equivalent satisfy the relation

$$z_a = \frac{z_{a1}\, z_{a2}}{z_{a1} + z_{a2}} + \frac{1}{z_{a1} + z_{a2}} = \frac{1}{z_b}.$$

Thus z_a of the resultant lattice is given by the series combination of z_{a1} and z_{a2} in parallel, and $1/z_{a1}$ and $1/z_{a2}$ in parallel; or alternatively by the parallel combination of z_{a1} and $1/z_{a2}$ in series, and z_{a2} and $1/z_{a1}$ in series. The inversion of this result serves as a means for decomposing a more complicated lattice into simpler components provided the proper relations are satisfied.

12–9. Recognizing the familiar network property that a parallel combination of Z and R, in *series* with a parallel combination of R^2/Z and R; or a series combination of Z and R, in *parallel* with the series combination of R^2/Z and R, is equivalent to the resistance R alone (Boucherôt network), show that if in the network arrangement of Problem 12–8 $z_{a2} = r$ (a pure resistance), we find for the resultant z_a *when* $r < 1$, a resistance r in series with the parallel combination of $(1 - r^2)/z_{a1}$ and $(1 - r^2)/r$, and *when* $r > 1$, a resistance $1/r$ in series with the parallel combination of $(r^2 - 1)/r^2 z_{a1}$ and $(r^2 - 1)/r$; or alternatively we find for the resultant z_a *when* $r < 1$, a resistance $1/r$ in parallel with the series combination of $1/z_{a1}(1 - r^2)$ and $r/(1 - r^2)$, and *when* $r > 1$, a resistance r in parallel with the series combination of $r^2 z_{a1}/(r^2 - 1)$ and $r/(r^2 - 1)$.

12-10. On the basis of the results of Problems 12-8 and 12-9 determine the conditions under which the lattice with a more complicated impedance z_a may be decomposed into lattices with components of the form shown in Fig. 214, or still simpler forms.

12-11. For the design of the phase corrective network in the text determine the bridged-T equivalent shown in Fig. 230*a*.

12-12. For the phase corrective network designed in the text find an equivalent cascade of simpler lattices and replace these by their bridged-T equivalents so that the final network will be in the form of a cascade of bridged-T networks.

12-13. Carry through the design of the phase corrective network discussed in the text by the alternative process of superposing the phase characteristics of simple component lattices, thus obtaining a cascaded form of structure directly. In order to facilitate this procedure, construct families of phase-versus-frequency characteristics for the simple lattices with z_a given by an inductance alone or by an inductance and capacitance in parallel, for a number of chosen parameter values. The process then becomes a graphical cut-and-try method of superposing a number of such elementary phase functions.

CHAPTER XIII

THE TRANSIENT BEHAVIOR OF LONG LINES

1. Introductory remarks. Since the various methods of analysis which were developed for lumped networks also apply to the determination of the terminal behavior of the electrically long line, the response of the latter as the result of the application of an arbitrary impressed force may evidently be treated in a number of ways. When the question at issue concerns itself with the sufficiency of a long line as the transmitting facility for a given form of applied signal, the Fourier integral method of reasoning discussed in Chapter XI is, for the communication engineer, the most satisfactory since it involves merely a study of the steady-state response functions and does not require that a specific transient analysis be carried out. When, for certain reasons, such a specific analysis is to be made, the Fourier integral method is impractical on account of the difficulty in carrying out the implied integrations by analytic means. In view of the fact that parameter variations with frequency may readily be taken into account in the Fourier method (which is not true of other methods), it is the only one capable of fully meeting the requirements of practical cases. When machine methods of integration become available, therefore, the Fourier method should stand out above all others with regard to the treatment of communication problems, but as an analytic method it must be displaced by other, more convenient, although less rigorous ones.

The word "convenient" must here be interpreted *cum grano salis*, for indeed, none of the methods for evaluating the transient behavior of a line with given terminal networks is convenient, even when certain sacrifices in the exact representation are made, such as neglecting parameter variations, neglecting or approximating the dissipative properties of the line and its terminal networks, or replacing the latter by approximate equivalents, and the like. Except in the simplest of approximate treatments (which are also readily treated analytically by the Fourier method), the various alternative processes of evaluation still involve quite a chore.

These alternative methods of solution may be classified as belonging to either of two general forms of treatment, to wit: the so-called solutions in terms of **wave functions** or in terms of **normal functions.** Of these the first is better adapted to the needs of the power engineer. It

does not give the solution in closed form valid for all values of time, but rather describes in mathematical language the wave picture on the line as it develops with increasing time. Although it can be extended to account for the resistance or for the net dissipative properties of the line, it is most readily applicable to the non-dissipative line with simple terminal networks (resistances only or a single non-dissipative coil or condenser, etc.). Reflection phenomena at the terminals or at junctions with other lines are treated singly at the time of their occurrence, and the development of the wave picture is followed in its increasing complexity (due to multiple reflections at ends, etc.) only so far as the requirements of the investigation demand. The solutions themselves may be got either by classical means[1] or by the use of the operational calculus,[2] the latter being particularly suitable for certain cases.

The solution in terms of normal functions, on the other hand, gives rise to a form valid for all values of time, and is in all respects the parallel of the classical method for the solution of lumped network problems. From this aspect the long-line problem represents an extension to a system having an infinite number of meshes (infinite number of degrees of freedom), and hence an infinite number of integration constants. The chief difficulty in the solution centers about the evaluation of these constants, which appear as coefficients in infinite series expansions. The number of practical cases which can conveniently be treated by this method is rather limited, but no more so than for solutions in terms of wave functions. The fact that the latter method of solution is usually looked upon as being more compact and readily applicable is (in the writer's opinion) largely due to the fact that more time and effort have recently been devoted toward further developing the detailed aspects of that method, whereas the normal function method of solution has received comparatively little attention in this respect since it was first introduced to the engineering profession by K. W. Wagner in 1908.[3]

[1] See, for example, "Elektrische Schaltvorgänge," R. Rüdenberg, Julius Springer, Berlin, 1923, p. 327, e.s., or "Electric Power Transmission and Distribution," L. F. Woodruff, John Wiley & Son, New York, 1925, pp. 251–284. See also V. Beweley, "Traveling Waves on Transmission Systems," John Wiley & Sons, 1933.

[2] See, for example, "Operational Circuit Analysis," V. Bush, John Wiley & Sons, New York, 1929, p. 189, e.s.; also J. R. Carson, "Electric Circuit Theory and Operational Calculus," McGraw-Hill, 1926, pp. 85–132. The student is also referred, of course, to the source of this material which is found in Heaviside's "Electromagnetic Theory," Vol. II. A further development of the subject is given in a Cambridge University tract entitled, "Operational Methods in Mathematical Physics," by Harold Jeffreys.

[3] "Elektromagnetische Ausgleichsvorgänge in Freileitungen und Kabeln," B. G. Teubner, 1908. Further developments may be found in E.T.Z., 34, 1913, p. 1053,

The treatment in terms of normal functions, furthermore, is the exact parallel of the classical method used in a large number of problems in physics and thermodynamics such as those dealing with the vibrations of strings and membranes, the flow of heat through homogeneous media, etc. Some familiarity with the philosophy of this method, therefore, is of general value with regard to vibration and diffusion problems; and a better acquaintance with the problem of expanding an arbitrary function in terms of sets of orthogonal normal functions, which is the point of chief interest in this method, is a valuable asset in the attack on many and diverse approximation problems met in lumped network design.[1]

The solution in terms of normal functions, moreover, is closely related to the Fourier integral form, reducing to this in the case of an infinitely long line or one terminated in its characteristic impedance. As a general method it is in some respects better adapted to the needs of the communication engineer than the solution in terms of wave functions. For these reasons, for its general mathematical value, and because a general discussion of its method of application is at present least available in the literature, we have chosen the method of solution in terms of normal functions for further particular discussion here.[2]

It may be mentioned in this connection that Heaviside's expansion formula bears the same relation to the normal function method of solution that it does to the classical solution of the lumped network problem; namely, it is an alternative method for evaluating the constants of integration.[3] In this respect it is more general in form, but in those cases which lend themselves to treatment at all, it does not represent a saving in time and effort. Therefore, we shall not consider the use of the Heaviside formula in our present discussion.

2. The solution in terms of normal functions. Our point of departure is from the partial differential equations representing the equilibrium

---

and in Archiv der Math. und Phys., Series III, Vol. 18, No. 3, pp. 230–241. Also by V. Bush, "Transmission Line Transients," A.I.E.E. Trans., 42, 1923, p. 878, and by C. L. Fortescue, A. L. Atherton, J. H. Cox, "Theoretical and Field Investigations of Lightning," A.I.E.E. Trans., 48, 1929, p. 449.

[1] The Cauer method of parameter determination (see section 7, Chapter X) or Bode's method of impedance correction (section 12, Chapter IX), for example, involve the mathematical aspects of this problem.

[2] Being the exact parallel of the lumped-constant treatment, the solution in normal functions naturally takes into account arbitrary initial distributions of voltage and current on the line, whereas with the wave solutions, by classical or operational methods, such distributions are taken into account by collateral modifications in treatment.

[3] See Heaviside's Vol. II; also Cohen's, "Heaviside Electric Circuit Theory," McGraw-Hill, 1928.

of a typical differential (internal) section of line (eq. (29) p. 33) with (assumed) constant parameters R, L, G, and C per unit length. These we repeat here for the sake of convenience

$$\left.\begin{aligned} \frac{\partial e}{\partial x} + L\frac{\partial i}{\partial t} + Ri = 0 \\ \frac{\partial i}{\partial x} + C\frac{\partial e}{\partial t} + Ge = 0 \end{aligned}\right\}. \tag{1058}$$

At present we are interested in obtaining the transient portion of the complete solution. As in the case of lumped networks we assume the forms

$$\left.\begin{aligned} e_\nu(x, t) &= \varphi_\nu(x) \cdot e^{p_\nu t} \\ i_\nu(x, t) &= \psi_\nu(x) \cdot e^{p_\nu t} \end{aligned}\right\}, \tag{1059}$$

which are termed **partial solutions** since they are written in terms of only one of the modes p_ν. In contrast to the lumped-network problem, the corresponding complex amplitudes are not constants but functions of distance x.

Substituting the assumptions (1059) into (1058) we obtain the condition equations

$$\left.\begin{aligned} (\varphi_\nu' + z_\nu\psi_\nu)e^{p_\nu t} = 0 \\ (\psi_\nu' + y_\nu\varphi_\nu)e^{p_\nu t} = 0 \end{aligned}\right\}, \tag{1060}$$

in which the prime denotes differentiation with respect to x, and the following symbols are introduced for abbreviation

$$\left.\begin{aligned} y_\nu &= G + Cp_\nu \\ z_\nu &= R + Lp_\nu \end{aligned}\right\}. \tag{1061}$$

For the existence of non-trivial solutions, the conditions (1060) evidently require that

$$\left.\begin{aligned} \varphi_\nu' + z_\nu\psi_\nu = 0 \\ \psi_\nu' + y_\nu\varphi_\nu = 0 \end{aligned}\right\}, \tag{1060a}$$

which are total differential equations in x alone.

By elimination we readily obtain from (1060a) the pair of equations in φ_ν and ψ_ν alone

$$\left.\begin{aligned} \varphi_\nu'' - \lambda_\nu^2\varphi_\nu = 0 \\ \psi_\nu'' - \lambda_\nu^2\psi_\nu = 0 \end{aligned}\right\}, \tag{1062}$$

in which we have let

$$\lambda_\nu^2 = y_\nu z_\nu. \tag{1063}$$

Using (1061), this gives

$$\lambda_\nu^2 = (R + Lp_\nu)(G + Cp_\nu); \tag{1063a}$$

or if we write

$$\left.\begin{aligned} \delta &= \frac{R}{2L} + \frac{G}{2C} \\ \sigma &= \frac{R}{2L} - \frac{G}{2C} \\ v_0^2 &= \frac{1}{LC} \end{aligned}\right\}, \tag{1064}$$

this is found equivalent to

$$v_0^2\lambda_\nu = (p_\nu + \delta)^2 - \sigma^2, \tag{1063b}$$

from which we have either

$$v_0\lambda_\nu \doteq (p_\nu + \delta)\left\{1 - \left(\frac{\sigma}{p_\nu + \delta}\right)^2\right\}^{\frac{1}{2}}, \tag{1065}$$

or

$$p_\nu = -\delta + v_0\lambda_\nu\left\{1 + \left(\frac{\sigma}{v_0\lambda_\nu}\right)^2\right\}^{\frac{1}{2}}. \tag{1065a}$$

From our familiarity with the steady-state solution we recognize that λ_ν is the propagation function for the mode p_ν. In mathematical physics the system of λ_ν's are known as the **proper values** or **eigenwerte** of the problem. They play the same rôle with regard to distance that the modes p_ν play with regard to time. The relations (1065) and (1065a) determine the λ_ν's in terms of the p_ν's, or vice versa. In most practical cases it is found that for almost all the modes

$$\left(\frac{\sigma}{p_\nu + \delta}\right)^2 \ll 1 \quad \text{or} \quad \left(\frac{\sigma}{v_0\lambda_\nu}\right)^2 \ll 1, \tag{1066}$$

so that we may use the approximate expressions

$$\left.\begin{aligned} v_0\lambda_\nu &\cong p_\nu + \delta \\ p_\nu &\cong -\delta + v_0\lambda_\nu \end{aligned}\right\}. \tag{1067}$$

Returning to equations (1062) we recognize that these are simultaneously satisfied by

$$\left.\begin{aligned} \varphi_\nu &= m_\nu e^{-\lambda_\nu x} + n_\nu e^{\lambda_\nu x} \\ \psi_\nu &= \frac{m_\nu e^{-\lambda_\nu x} - n_\nu e^{\lambda_\nu x}}{Z_{0\nu}} \end{aligned}\right\}, \tag{1068}$$

where

$$Z_{0\nu} = \sqrt{\frac{z_\nu}{y_\nu}} = \sqrt{\frac{R + Lp_\nu}{G + Cp_\nu}} \tag{1069}$$

is used to denote the characteristic impedance for the mode p_ν, and m_ν and n_ν are constants to be determined from the boundary conditions. For this purpose we refer to Fig. 233 which represents a single line diagram for the line with terminal networks whose impedances to the mode p_ν are denoted by Z_{S_ν} and Z_{R_ν} for the sending ($x = 0$) and receiving ($x = l$) ends, respectively. Since we are discussing the transient (force-free) solution, no impressed emf appears in this picture.

The boundary conditions for the partial voltages φ_ν and currents ψ_ν evidently require that

$$\left.\begin{aligned} \varphi_\nu(0) + Z_{S\nu}\psi_\nu(0) &= 0 \\ \varphi_\nu(l) - Z_{R\nu}\psi_\nu(l) &= 0 \end{aligned}\right\}. \tag{1070}$$

Substituting into these from (1068) we find

$$\left.\begin{aligned} (Z_{0\nu} + Z_{S\nu})m_\nu + \quad (Z_{0\nu} - Z_{S\nu})n_\nu &= 0 \\ (Z_{0\nu} - Z_{R\nu})e^{-\lambda_\nu l}m_\nu + (Z_{0\nu} + Z_{R\nu})e^{\lambda_\nu l}n_\nu &= 0 \end{aligned}\right\}. \tag{1070a}$$

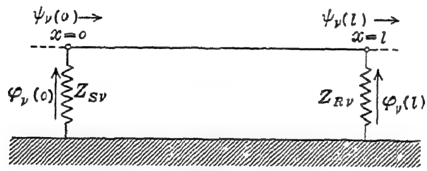


FIG. 233.—Single-line schematic diagram showing the boundary conditions for the transient behavior of the long line

This pair of homogeneous equations has non-zero solutions for m_ν and n_ν only when its determinant vanishes. The vanishing of this determinant constitutes the determinantal equation for this problem; i.e., it is that equation from which the modes p_ν or the eigenwerte λ_ν are found for a given set of terminal networks at the line ends. This determinantal equation may be written in various alternative forms, for example

$$\left.\begin{aligned} &(Z_{S\nu}Z_{R\nu} + Z_{0\nu}^2) \sinh \lambda_\nu l + Z_{0\nu}(Z_{S\nu} + Z_{R\nu}) \cosh \lambda_\nu l = 0 \\ \text{or} \qquad &\frac{Z_{0\nu}(Z_{S\nu} + Z_{R\nu})}{Z_{S\nu}Z_{R\nu} + Z_{0\nu}^2} + \tanh \lambda_\nu l = 0 \end{aligned}\right\}. \tag{1071}$$

Writing for the **partial reflection coefficients**

$$r_{S\nu} = \frac{Z_{S\nu} - Z_{0\nu}}{Z_{S\nu} + Z_{0\nu}}; \; r_{R\nu} = \frac{Z_{R\nu} - Z_{0\nu}}{Z_{R\nu} + Z_{0\nu}}, \tag{1072}$$

we obtain a very compact form, namely

$$r_{S\nu}r_{R\nu} = e^{2\lambda_\nu l}. \tag{1071a}$$

For the further development of the solution it is convenient to write

$$\left.\begin{aligned} r_{S_\nu} &= e^{2\rho_{S_\nu}};\ r_{R_\nu} = e^{2\rho_{R_\nu}} \\ \rho_{S_\nu} &= \tfrac{1}{2}\ln r_{S_\nu};\ \rho_{R_\nu} = \tfrac{1}{2}\ln r_{R_\nu} \end{aligned}\right\}, \tag{1073}$$

by means of which the determinantal equation takes the form

$$\rho_{S_\nu} + \rho_{R_\nu} = \lambda_\nu l. \tag{1071b}$$

Since Z_{S_ν} and Z_{R_ν} (or the equivalent pairs r_{S_ν}, r_{R_ν}, or ρ_{S_ν}, ρ_{R_ν}) are functions of p_ν, and the relation between p_ν and λ_ν, according to (1063a), involves a radical, the determinantal equation in the general case is difficult to solve, particularly since it contains a transcendental function of λ_ν. On this latter account the number of modes (or eigenwerte) is infinite, as we expected. In general the determinantal equation is solved graphically.[1] A discussion of this for several specific cases is taken up later.

The vanishing of the determinant of (1070a) determines the following ratio of m_ν to n_ν

$$\frac{m_\nu}{n_\nu} = r_{S_\nu} = \frac{e^{2\lambda_\nu l}}{r_{R_\nu}}. \tag{1074}$$

Except for an arbitrary coefficient we may, with the help of (1073), write the functions (1068) in the following form

$$\left.\begin{aligned} \varphi_\nu(x) &= \sqrt{\frac{2}{l}}\cosh(\lambda_\nu x - \rho_{S_\nu}) \\ \psi_\nu(x) &= -\frac{1}{Z_{0_\nu}}\sqrt{\frac{2}{l}}\sinh(\lambda_\nu x - \rho_{S_\nu}) \end{aligned}\right\}. \tag{1075}$$

The factor $\sqrt{2/l}$ is introduced for reasons which will become clear in section 7 below.

Denoting the arbitrary coefficients (transient current amplitudes) by A_ν, the complete transient solutions for voltage and current become

$$\left.\begin{aligned} e_t(x, t) &= \sum_{\nu=-\infty}^{\infty} A_\nu \varphi_\nu(x) e^{p_\nu t} \\ i_t(x, t) &= \sum_{\nu=-\infty}^{\infty} A_\nu \psi_\nu(x) e^{p_\nu t} \end{aligned}\right\}. \tag{1076}$$

Since the functions (1075) satisfy the simultaneous equations (1060a), the coefficients A_ν are the same in the voltage and current solutions.

The functions (1075) are known as the **normal functions** of our problem. In general they are complex, as are the amplitudes A_ν.

[1] Provided a solution is at all possible, of course. The necessity for determining the modes of the system is one of the weak points of this method of solution which is obviated by the treatment in terms of wave functions.

Owing to the infinite number of modes, the solutions (1076) contain an infinite number of terms and hence an infinite number of transient amplitudes A_ν. These must now be determined from the initial conditions which are the initial voltage (charge) and current distributions on the line. This process is formulated more definitely as follows.

The complete voltage and current solutions are given by the sum of the transient and steady-state parts. Denoting the latter by means of the subscript s we have

$$\left.\begin{aligned} e(x, t) &= e_s(x, t) + e_t(x, t) \\ i(x, t) &= i_s(x, t) + i_t(x, t) \end{aligned}\right\}. \tag{1077}$$

For the initial instant, $t = 0$, the net voltage and current $e(x, 0)$ and $i(x, 0)$ may be arbitrarily given functions of x. For the initial *transient* distributions of voltage and current, therefore, we have

$$\left.\begin{aligned} P(x) &= e_t(x, 0) = e(x, 0) - e_s(x, 0) \\ Q(x) &= i_t(x, 0) = i(x, 0) - i_s(x, 0) \end{aligned}\right\}. \tag{1077a}$$

Using these in (1076) we find that the following relations must be satisfied

$$\left.\begin{aligned} P(x) &= \sum_{\nu=-\infty}^{\infty} A_\nu \varphi_\nu(x) \\ Q(x) &= \sum_{\nu=-\infty}^{\infty} A_\nu \psi_\nu(x) \end{aligned}\right\}. \tag{1078}$$

These represent infinite expansions of the functions $P(x)$ and $Q(x)$ in terms of the normal functions $\varphi_\nu(x)$ and $\psi_\nu(x)$. The problem here is to determine the coefficients A_ν so as to satisfy (1078) over the region $0 < x < l$ for which the arbitrary distributions are specified. The reader should note that, as we have written them, the summations extend over all positive and negative integer values of ν. We shall dispose of the numbering of the modes p_ν and of the eigenwerte λ_ν in such a way that corresponding positive and negative integers refer to conjugate complex values and hence to conjugate complex values for the functions $\varphi_\nu(x)$ and $\psi_\nu(x)$ as well as for the coefficients A_ν also. Since the terms in the expansions (1078) thus come in conjugate pairs, the resulting summations are real functions of x as they should be. It is possible, of course, to combine such conjugate terms and write the expansions (1078) with real terms only. For certain simple cases this simplifies matters somewhat, but for the discussion of the general case the complex forms are more convenient and, therefore, we shall leave them so.

3. The expansion problem. We come now to the chief difficulty in the solution to our problem, namely, that of determining the coefficients

in the expansions (1078). The complete discussion of this collateral problem is exceedingly extensive and occupies a very prominent position in the field of mathematical physics. For the engineer, however, only such cases are of interest for which the solution may be carried out without an undue amount of labor; otherwise alternative methods of solution or reasonable approximations are resorted to rather than attempt to meet the situation on the present basis. We shall discuss here only those methods for the evaluation of the coefficients which are readily carried out, and attempt to classify those boundary conditions (types of terminal networks) to which these simpler methods apply.

The simplest method for attempting an evaluation of the coefficients A_ν is that originally suggested by Fourier. When φ_ν or ψ_ν are trigonometric functions with arguments which are integer multiples of one of these, this leads to the familiar *Fourier series expansions*. The process is usually outlined as follows:

If we multiply the first equation (1078) by $\varphi_\mu(x)$, the second by $\psi_\mu(x)$, μ being any other integer than ν, and integrate the results over the fundamental region $0 < x < l$, we get

$$\left.\begin{aligned} \int_0^l P(x)\varphi_\mu(x)dx &= \sum_{\nu=-\infty}^{\infty} A_\nu \int_0^l \varphi_\nu\varphi_\mu dx \\ \int_0^l Q(x)\psi_\mu(x)dx &= \sum_{\nu=-\infty}^{\infty} A_\nu \int_0^l \psi_\nu\psi_\mu dx \end{aligned}\right\}. \tag{1079}$$

Assuming that the following integrals vanish for $\nu \neq \pm\mu$, i.e., that

$$\left.\begin{aligned} \int_0^l \varphi_\nu\varphi_\mu dx &= 0 \\ \int_0^l \psi_\nu\psi_\mu dx &= 0 \end{aligned}\right\}(\nu \neq \pm\mu), \tag{1080}$$

then the infinite summations on the right of (1079) reduce to only two terms, namely, that for $\nu = \mu$ and the conjugate term for $\nu = -\mu$. This gives

$$\left.\begin{aligned} \int_0^l P(x)\varphi_\mu(x)dx &= A_\mu\int_0^l \varphi_\mu^2 dx + A_{-\mu}\int_0^l |\varphi_\mu|^2 dx \\ \int_0^l Q(x)\psi_\mu(x)dx &= A_\mu\int_0^l \psi_\mu^2 dx + A_{-\mu}\int_0^l |\psi_\mu|^2 dx \end{aligned}\right\}. \tag{1081}$$

Regarding A_μ and $A_{-\mu}$ as unknowns, this pair of algebraic equations is solvable provided its determinant does not vanish. In the singular case, i.e., if the determinant does vanish, the physical conditions are evidently such that only the voltage distribution $P(x)$ or the current distribution $Q(x)$ may be arbitrarily specified *but not both independently*.

In that case the functions φ_ν and ψ_ν become real, as do the coefficients A_ν, and the problem simplifies as compared to the more general formulation (1081).

When the conditions (1080) are satisfied the functions φ_ν and ψ_ν are said to be **orthogonal.** This term is evidently taken from the analogous situation in geometry where a pair of vectors are orthogonal if their scalar product vanishes. The integrals (1080) are the analogues (in terms of the continuous variable x) of such scalar products. In the orthogonal case the coefficients A_ν are quite readily determined. Note, however, that *both* functions φ_ν *and* ψ_ν must be orthogonal. The terminal conditions which lead to simultaneous orthogonality for φ_ν and ψ_ν are relatively restricted. This will be discussed in more detail below. At present we wish to investigate whether simple modifications of this method may still lead to solutions in case the relations (1080) are not fulfilled.

A rather obvious suggestion, if the method just outlined fails, is to try a linear combination of the integrals (1079). If q be an arbitrary factor, then let us form

$$\int_0^l (P(x)\varphi_\mu + qQ(x)\psi_\mu)dx = \sum_{\nu=-\infty}^{\infty} A_\nu \int_0^l (\varphi_\nu\varphi_\mu + q\psi_\nu\psi_\mu)dx. \quad (1082)$$

If we now find that

$$\int_0^l (\varphi_\nu\varphi_\mu + q\psi_\nu\psi_\mu)dx = 0; \; (\nu \neq \pm \mu), \quad (1083)$$

and, further, if in addition *either*

$$\int_0^l (\varphi_\mu^2 + q\psi_\mu^2)dx = 0 \quad (1083a)$$

or

$$\int_0^l (|\varphi_\mu|^2 + q|\psi_\mu|^2)dx = 0, \quad (1083b)$$

then the coefficients may evidently be found respectively from *either*

$$A_\mu = \frac{\displaystyle\int_0^l (P(x)\overline{\varphi}_\mu + qQ(x)\overline{\psi}_\mu)dx}{\displaystyle\int_0^l (|\varphi_\mu|^2 + q|\psi_\mu|^2)dx}, \quad (1082a)$$

or

$$A_\mu = \frac{\displaystyle\int_0^l (P(x)\varphi_\mu + qQ(x)\psi_\mu)dx}{\displaystyle\int_0^l (\varphi_\mu^2 + q\psi_\mu^2)dx}, \quad (1082b)$$

where the bar indicates the conjugate value. Although the conditions for this process of evaluation seem to be more rigid than the former, we shall see that they admit the treatment of a number of additional cases.

Further latitude is gained in the variety of cases which may be treated on the Fourier basis if we assume that the functions $P(x)$ and $Q(x)$ as well as their series expansions are differentiable, i.e., if we can write

$$\left.\begin{aligned} P'(x) &= \sum_{\nu=-\infty}^{\infty} A_\nu \varphi_\nu'(x) \\ Q'(x) &= \sum_{\nu=-\infty}^{\infty} A_\nu \psi_\nu'(x) \end{aligned}\right\}. \tag{1084}$$

Although (as we shall see) the conditions for the simultaneous orthogonality of φ_ν' and ψ_ν' are the same as those for φ_ν and ψ_ν, those for φ_ν and ψ_ν', or φ_ν' and ψ_ν are different and admit of the treatment of additional cases.

By means of manipulations similar to these, the Fourier method may be considerably extended. The above detailed discussion should give the reader a sufficient idea of the principles involved in order that he may continue along these lines by himself. When the Fourier method fails, certain additional cases may be treated with a slight increase in labor by a process originated by Lord Rayleigh[1] which we shall discuss briefly later. At present we shall consider in some detail the boundary conditions which lead to one or the other of the forms of orthogonality mentioned above.

4. Terminal conditions leading to orthogonality. Here equations (1062) are taken as a point of departure. Consider the first of these multiplied by φ_μ, and the resulting equation written again with the subscripts ν and μ interchanged, thus

$$\left.\begin{aligned} \varphi_\nu''\varphi_\mu - \lambda_\nu^2\varphi_\nu\varphi_\mu &= 0 \\ \varphi_\mu''\varphi_\nu - \lambda_\mu^2\varphi_\mu\varphi_\nu &= 0 \end{aligned}\right\}. \tag{1085}$$

Subtracting and integrating over the line length gives

$$(\lambda_\nu^2 - \lambda_\mu^2)\int_0^l \varphi_\nu\varphi_\mu dx = \int_0^l (\varphi_\nu''\varphi_\mu - \varphi_\mu''\varphi_\nu)dx.$$

Integrating the right-hand side by parts, we find

$$(\lambda_\nu^2 - \lambda_\mu^2)\int_0^l \varphi_\nu\varphi_\mu dx = \Big[\varphi_\mu\varphi_\nu' - \varphi_\nu\varphi_\mu'\Big]_0^l. \tag{1086}$$

By a similar process we get from the second equation (1062)

$$(\lambda_\nu^2 - \lambda_\mu^2)\int_0^l \psi_\nu\psi_\mu dx = \Big[\psi_\mu\psi_\nu' - \psi_\nu\psi_\mu'\Big]_0^l. \tag{1087}$$

[1] "The Theory of Sound," Vol. I, p. 200 e.s. See also K. W. Wagner, *loc. cit.* pp. 71–73.

Simultaneous orthogonality of φ_ν and ψ_ν according to (1080) results if the right-hand sides of (1086) and (1087) vanish. Orthogonality of a linear combination according to (1083) results when corresponding linear combinations of the right-hand sides of (1086) and (1087) vanish. This we wish to express directly in terms of the terminal impedances and the line parameters. To this end we first combine the differential equations (1060a) with the boundary conditions (1070) to form the following relations

$$\left.\begin{aligned} \varphi_\nu'(0) &= \frac{z_\nu}{Z_{S\nu}}\varphi_\nu(0);\ \varphi_\nu'(l) = -\frac{z_\nu}{Z_{R\nu}}\varphi_\nu(l) \\ \psi_\nu'(0) &= y_\nu Z_{S\nu}\psi_\nu(0);\ \psi_\nu'(l) = -y_\nu Z_{R\nu}\psi_\nu(l) \end{aligned}\right\}. \tag{1088}$$

Substitution of these into (1086) and (1087) gives

$$(\lambda_\nu^2 - \lambda_\mu^2)\int_0^l \varphi_\nu\varphi_\mu dx = \varphi_\nu(0)\varphi_\mu(0)\left(\frac{z_\mu}{Z_{S\mu}} - \frac{z_\nu}{Z_{S\nu}}\right) + \varphi_\nu(l)\varphi_\mu(l)\left(\frac{z_\mu}{Z_{R\mu}} - \frac{z_\nu}{Z_{R\nu}}\right), \tag{1089}$$

and

$$(\lambda_\nu^2-\lambda_\mu^2)\int_0^l \psi_\nu\psi_\mu dx = \psi_\nu(0)\psi_\mu(0)(y_\mu Z_{S\mu} - y_\nu Z_{S\nu}) + \psi_\nu(l)\psi_\mu(l)(y_\mu Z_{R\mu} - y_\nu Z_{R\nu}). \tag{1090}$$

Equations (1060a) further show that in these relations we may replace φ by ψ' and ψ by φ'. Hence the corresponding orthogonality conditions for φ' and ψ' are immediately evident.

It is thus easy to see from equation (1089) that the functions $\varphi_\nu(x)$ and $\psi_\nu'(x)$ become orthogonal if

$$\left.\begin{aligned} Z_{S\nu} &= az_\nu = a(R + Lp_\nu) \\ Z_{R\nu} &= bz_\nu = b(R + Lp_\nu) \end{aligned}\right\}, \tag{1091}$$

where a and b are any constants inclusive of zero and infinity (short-circuited and open-circuited line ends, respectively). For finite values of a and b this represents series combinations of resistance and inductance in the same ratio as that for the line.

Equation (1090), on the other hand, shows that $\varphi_\nu'(x)$ and $\psi_\nu(x)$ become orthogonal if

$$\left.\begin{aligned} \frac{1}{Z_{S\nu}} &= a'y_\nu = a'(G + Cp_\nu) \\ \frac{1}{Z_{R\nu}} &= b'y_\nu = b'(G + Cp_\nu) \end{aligned}\right\}, \tag{1092}$$

where a' and b' are again any constants inclusive of the values zero and infinity. This represents terminal impedances which are parallel com-

binations of conductance and capacitance in the same ratio as that occurring on the line.

A further study of (1089) and (1090), using the relations (1072), (1073), and (1075), reveals that both $\varphi_\nu(x)$ and $\psi_\nu(x)$ become orthogonal if

$$\left.\begin{aligned} Z_{S\nu} &= dZ_{0\nu} \\ Z_{R\nu} &= \frac{Z_{0\nu}}{d} \end{aligned}\right\}, \tag{1093}$$

where d is any constant inclusive of zero and infinity (one end short-circuited and the other open) but different from unity. The latter value corresponds to proper termination and hence to a virtually infinitely long line. This case, which requires special treatment, will be discussed later.

A study of these same relations shows that orthogonality for the linear combination of $\varphi_\nu(x)$ and $\psi_\nu(x)$ according to (1083) results for either of two cases, both of which require that $Z_{0\nu} = Z_0 =$ constant. One of these, which gives rise to (1083*a*) and hence leads to the evaluation of A_μ by means of (1082*a*), is further defined by

$$\left.\begin{aligned} Z_{S\nu} &= a_\nu Z_0 \\ Z_{R\nu} &= \frac{Z_0}{a_\nu} \\ q &= Z_0^2 \end{aligned}\right\}. \tag{1094}$$

The other, which gives rise to (1083*b*) and hence to the evaluation of A_μ by means of (1082*b*), occurs for

$$\left.\begin{aligned} Z_S &= aZ_0 \\ Z_R &= bZ_0 \\ q &= -Z_0^2 \end{aligned}\right\}. \tag{1095}$$

In (1094) a_ν is any function of p_ν. This case enables one to treat by the Fourier method any combination of *inverse* terminal networks whose impedance product is Z_0^2. In (1095) a and b are any constants different from unity. This case represents any combination of terminal resistances differing from Z_0, and is thus an extension of (1093), although the latter does not require that Z_0 be a constant. This requirement is met only by the non-dissipative or the non-distortive line for which $R/L = G/C$.

These cases by no means exhaust all possible combinations of terminal networks although they represent an extension over the usual ones treated by the Fourier method. Further development of the process sketched here may lead to other useful results. When this fails, recourse may be had to another method which we shall discuss briefly.

5. Terminal conditions to which Rayleigh's method applies. Suppose that the terminal impedances can be represented as

$$\left.\begin{aligned} \frac{1}{Z_{S\nu}} &= a_1 y_\nu + \frac{1}{a_2 z_\nu} \\ \frac{1}{Z_{R\nu}} &= b_1 y_\nu + \frac{1}{b_2 z_\nu} \end{aligned}\right\}, \tag{1096}$$

i.e., as parallel combinations of impedances which are proportional to $1/y_\nu$ and z_ν. The factors a_1, a_2, b_1, and b_2 may be any constants. By means of (1063) we then have

$$\frac{z_\nu}{Z_{S\nu}} = a_1\lambda_\nu^2 + \frac{1}{a_2},$$

and

$$\frac{z_\nu}{Z_{R\nu}} = b_1\lambda_\nu^2 + \frac{1}{b_2},$$

from which we get

$$\frac{z_\mu}{Z_{S\mu}} - \frac{z_\nu}{Z_{S\nu}} = -\,a_1(\lambda_\nu^2 - \lambda_\mu^2), \tag{1097}$$

and

$$\frac{z_\mu}{Z_{R\mu}} - \frac{z_\nu}{Z_{R\nu}} = -\,b_1(\lambda_\nu^2 - \lambda_\mu^2). \tag{1097a}$$

By means of (1060a) and (1089) we then have for $\nu \neq \pm\,\mu$

$$\int^l \varphi_\nu\varphi_\mu dx = -\,a_1\varphi_\nu(0)\varphi_\mu(0) - b_1\varphi_\nu(l)\varphi_\mu(l), \tag{1098}$$

and

$$\int_0^l \psi_\nu'\psi_\mu' dx = -\,a_1\psi_\nu'(0)\psi_\mu'(0) - b_1\psi_\nu'(l)\psi_\mu'(l). \tag{1098a}$$

If, on the other hand, the terminal impedances may be represented as

$$\left.\begin{aligned} Z_{S\nu} &= a_1 z_\nu + \frac{1}{a_2 y_\nu} \\ Z_{R\nu} &= b_1 z_\nu + \frac{1}{b_2 y_\nu} \end{aligned}\right\}, \tag{1099}$$

i.e., as series combinations of impedances which are proportional to $1/y_\nu$ and z_ν, then we have

$$y_\nu Z_{S\nu} = a_1\lambda_\nu^2 + \frac{1}{a_2},$$

$$y_\nu Z_{R\nu} = b_1\lambda_\nu^2 + \frac{1}{b_2},$$

so that

$$y_\mu Z_{S\mu} - y_\nu Z_{S\nu} = -a_1(\lambda_\nu^2 - \lambda_\mu^2), \tag{1100}$$

and

$$y_\mu Z_{R\mu} - y_\nu Z_{R\nu} = -b_1(\lambda_\nu^2 - \lambda_\mu^2). \tag{1100a}$$

By means of (1060a) and (1090) we have for $\nu \neq \pm\mu$

$$\int_0^l \psi_\nu\psi_\mu dx = -a_1\psi_\nu(0)\psi_\mu(0) - b_1\psi_\nu(l)\psi_\mu(l), \tag{1101}$$

and

$$\int_0^l \varphi_\nu'\varphi_\mu' dx = -a_1\varphi_\nu'(0)\varphi_\mu'(0) - b_1\varphi_\nu'(l)\varphi_\mu'(l). \tag{1101a}$$

The results (1098) and (1098a) or (1101) and (1101a) enable us to evaluate the coefficients A_μ for the corresponding cases as follows.

In the first equation (1079) substitute (1098) and obtain

$$\begin{aligned}\int_0^l P(x)\varphi_\mu dx = &-\sum_{\nu=-\infty}^{\infty} A_\nu\{a_1\varphi_\nu(0)\varphi_\mu(0) + b_1\varphi_\nu(l)\varphi_\mu(l)\}\\ &+ A_\mu\left\{\int_0^l \varphi_\mu^2 dx + a_1\varphi_\mu^2(0) + b_1\varphi_\mu^2(l)\right\}\\ &+ A_{-\mu}\left\{\int_0^l |\varphi_\mu|^2 dx + a_1|\varphi_\mu(0)|^2 + b_1|\varphi_\mu(l)|^2\right\}.\end{aligned} \tag{1102}$$

Since the sum includes the terms for $\nu = \pm\mu$, for which (1098) does not hold, these terms are correctly added in the bracketed terms and the erroneous ones subtracted.

Now we see that

$$\sum_{\nu=-\infty}^{\infty} A_\nu a_1\varphi_\nu(0)\varphi_\mu(0) = a_1\varphi_\mu(0)P(0), \tag{1103}$$

and

$$\sum_{\nu=-\infty}^{\infty} A_\nu b_1\varphi_\nu(l)\varphi_\mu(l) = b_1\varphi_\mu(l)P(l). \tag{1104}$$

Hence (1102) becomes

$$\begin{aligned}\int_0^l P(x)\varphi_\mu dx + a_1\varphi_\mu(0)P(0) + b_1\varphi_\mu(l)P(l) = \\ A_\mu\left\{\int_0^l \varphi_\mu^2 dx + a_1\varphi_\mu^2(0) + b_1\varphi_\mu^2(l)\right\}\\ + A_{-\mu}\left\{\int_0^l |\varphi_\mu|^2 dx + a_1|\varphi_\mu(0)|^2 + b_1|\varphi_\mu(l)|^2\right\}.\end{aligned} \tag{1105}$$

In an analogous fashion we obtain from an equation similar to the

second of the pair (1079) but formed in terms of the differentiated functions, together with (1098a)

$$\int_0^l Q'(x)\psi_\mu' dx + a_1\psi_\mu'(0)Q'(0) + b_1\psi_\mu'(l)Q'(l) =$$
$$A_\mu\left\{\int_0^l \psi_\mu'^2 dx + a_1\psi_\mu'^2(0) + b_1\psi_\mu'^2(l)\right\}$$
$$+ A_{-\mu}\left\{\int_0^l |\psi_\mu'|^2 dx + a_1|\psi_\mu'(0)|^2 + b_1|\psi_\mu'(l)|^2\right\}. \quad (1106)$$

Thus we see that terminal impedances which satisfy (1096) lead to the pair of simultaneous equations (1105) and (1106), which may be solved for the coefficients A_μ provided their determinant does not vanish and provided the derivative $Q'(x)$ exists.

When the terminal impedances satisfy (1099), then a similar treatment utilizing (1101) and (1101a) leads to a result of the form (1105) with $P(x)$ replaced by $Q(x)$ and φ_μ by ψ_μ, and to a result of the form (1106) with $Q'(x)$ replaced by $P'(x)$ and ψ_μ' by φ_μ'. This leads to solutions for A_μ provided the derivative $P'(x)$ exists.

K. W. Wagner[1] gives an extension of the Rayleigh method to the non-dissipative line terminated in arbitrary non-dissipative networks (applicable also to the case for which the system as a whole is uniformly dissipative). The problem, however, is not of sufficient practical importance to carry the unusually complicated discussion any further here.

6. Solution of the determinantal equation. Although the determination of the λ_ν's or p_ν's is in general very involved, the process becomes quite simple for the special examples discussed above. Since the impedances (1096) and (1099) contain (1091) and (1092) as special cases, these may be treated together. For (1096), we get with the help of (1063) and (1069)

$$\left.\begin{aligned} r_{S\nu} &= -\frac{a_1a_2\lambda_\nu^2 - a_2\lambda_\nu + 1}{a_1a_2\lambda_\nu^2 + a_2\lambda_\nu + 1} \\ r_{R\nu} &= -\frac{b_1b_2\lambda_\nu^2 - b_2\lambda_\nu + 1}{b_1b_2\lambda_\nu^2 + b_2\lambda_\nu + 1} \end{aligned}\right\}, \quad (1107)$$

while for the impedances (1099) the same functions are obtained except for a change of sign. The determinantal equation according to (1071a), therefore, is the same for both cases.

From the form of the functions (1107) we recognize that the λ_ν's must

[1] Archiv der Math. und Phys., Ser. 3, Vol. 18, No. 3, pp. 230–241.

be pure imaginary numbers. Whereas in the general case we may write

$$\left.\begin{aligned} \lambda_\nu &= \gamma_\nu + j\omega_\nu \\ \rho &= \varepsilon_\nu + j\vartheta_\nu \end{aligned}\right\}, \tag{1108}$$

where the latter may refer to either reflection coefficient according to (1073), we have in the cases (1096) and (1099) $\lambda_\nu = j\omega_\nu$; $\rho_{S\nu} = j\vartheta_{S\nu}$ $\rho_{R\nu} = j\vartheta_{R\nu}$. The eigenwerte are pure imaginary quantities, and the reflection coefficients are unit vectors. In particular we have for (1096), according to (1107),

$$\left.\begin{aligned} \vartheta_{S\nu} &= \frac{\pi}{2} + \tan^{-1}\left(\frac{a_2\omega_\nu}{a_1a_2\omega_\nu^2 - 1}\right) \\ \vartheta_{R\nu} &= \frac{\pi}{2} + \tan^{-1}\left(\frac{b_2\omega_\nu}{b_1b_2\omega_\nu^2 - 1}\right) \end{aligned}\right\}, \tag{1107a}$$

while for (1099) the same values without the $\pi/2$ terms apply.

Utilizing (1071*b*) and combining the two anti-tangents, the determinantal equation for these cases is given by

$$\frac{H\omega_\nu(\omega_\nu^2 - \omega_1^2)}{(\omega_\nu^2 - \omega_2^2)(\omega_\nu^2 - \omega_4^2)} = \tan \omega_\nu l, \tag{1109}$$

where

$$H = \frac{a_1 + b_1}{a_1b_1},$$

$$\omega_1^2 = \frac{a_2 + b_2}{a_2b_2(a_1 + b_1)},$$

$$\omega_{2,4}^2 = \frac{m \mp \sqrt{m^2 - 4n}}{2n},$$

$$m = a_1a_2 + b_1b_2 + a_2b_2,$$

$$n = a_1a_2b_1b_2.$$

The function on the left-hand side of (1109) has the same form as a reactance except that its slope is everywhere negative. The intersections of this function with the tangent determine the ω_ν's and hence the p_ν's also. For the special cases (1091) and (1092), obvious simplifications result.

The cases (1093) and (1095) for pure resistance terminations in combination with either a dissipationless or a distortionless line may also be treated together. Since (1093) is a special case of (1095), we shall confine our attention to the latter. Here the reflection coefficients r_S and r_R are real and independent of the eigenwerte λ_ν. Utilizing

(1108), the determinantal equation according to (1071*b*) and (1073) becomes

$$\left.\begin{aligned}(\gamma_\nu + j\omega_\nu)l &= (\varepsilon_{S\nu} + \varepsilon_{R\nu}) + j(\vartheta_{S\nu} + \vartheta_{R\nu}) \\ &= \tfrac{1}{2}\ln r_S r_R\end{aligned}\right\}. \tag{1110}$$

The real parts γ_ν as well as ε_{S_ν} and ε_{R_ν} are evidently constant and given by

$$\gamma l = \varepsilon_S + \varepsilon_R = \tfrac{1}{2}\ln |r_S r_R|, \tag{1111}$$

while the imaginary parts are given by either

$$\omega_\nu l = \vartheta_{S\nu} + \vartheta_{R\nu} = \pi\nu;\ (\nu = \pm \text{ any integer}), \tag{1111a}$$

or

$$\omega_\nu l = \vartheta_{S\nu} + \vartheta_{R\nu} = \frac{\pi\nu}{2};\ (\nu = \pm \text{ odd integers}), \tag{1111b}$$

according to whether $r_S r_R$ is greater or less than zero, respectively. The former occurs when both terminal resistances are either larger or smaller than Z_0, and the latter when one is larger and the other smaller. When the line ends are open- or short-circuited γl, ε_S and ε_R, are zero, while the imaginary parts are given by (1111*a*) or (1111*b*). Since the magnitude of $r_S r_R$ in this case is always less than unity (the case $r_S r_R = 1$ is excluded), the real parts according to (1111) are always negative. When $r_S r_R < 0$, as it always is in the special case (1093), $\vartheta_{S_\nu} - \vartheta_{R_\nu} = \pm\, \pi/2$; for (1093) we also have $\varepsilon_S - \varepsilon_R = 0$.

For the remaining case (1094) the reflection coefficients are given by

$$r_{S_\nu} = -\, r_{R_\nu} = \frac{a_\nu - 1}{a_\nu + 1}, \tag{1112}$$

so that the determinantal equation becomes

$$\frac{a_\nu - 1}{a_\nu + 1} = e^{\lambda_\nu l \pm j\frac{\pi}{2}},$$

which is equivalent to

$$a_\nu = -\coth\left(\frac{\lambda_\nu l}{2} \pm j\frac{\pi}{4}\right). \tag{1113}$$

In general, a_ν may be the impedance function (in terms of p_ν) of a perfectly arbitrary network. Then no solution can be obtained by simple means. When the system as a whole may be assumed uniformly dissipative, then the network having a_ν as its driving-point impedance consists of coils with impedances $\alpha_k z_\nu$ and condensers with admittances

$\alpha_k y_\nu$, where the α_k's are any positive constants. The impedance function of such a network of n meshes is given by

$$a_\nu Z_0 = Z_0\left\{\alpha_1\lambda_\nu + \frac{1}{\alpha_2\lambda_\nu} + \frac{\alpha_3\lambda_\nu}{\alpha_3\alpha_4\lambda_\nu^2 + 1} + \cdots + \frac{\alpha_{2n-1}\lambda_\nu}{\alpha_{2n-1}\alpha_{2n}\lambda_\nu^2 + 1}\right\}. \tag{1114}$$

From the form of this expression we see that (1113) is satisfied by pure imaginary values of λ_ν. Writing (1114) in its factored form, (1113) becomes

$$\frac{H(\omega_\nu^2 - \omega_1^2)(\omega_\nu^2 - \omega_3^2)\cdots(\omega_\nu^2 - \omega^2{}_{2n-1})}{\omega_\nu(\omega_\nu^2 - \omega_2^2)(\omega_\nu^2 - \omega_4^2)\cdots(\omega_\nu^2 - \omega^2{}_{2n-2})} = \cot\left(\frac{\omega_\nu l}{2} \pm \frac{\pi}{4}\right). \tag{1113a}$$

Intersections of plots of these functions give the desired ω_ν-values. It is important to note that two cotangent curves are to be drawn, one for the phase $+\pi/4$ and one for $-\pi/4$.

7. Evaluation of integrals. In order to facilitate the evaluation of the coefficients A_ν in any of the above examples, it is useful to have explicit expressions for the integrals

$$\left.\begin{aligned} J_1 &= \int_0^l \varphi_\nu^2 dx;\ J_2 = \int_0^l |\varphi_\nu|^2 dx \\ J_3 &= Z_{0\nu}^{\ 2}\int_0^l \psi_\nu^2 dx;\ J_4 = |Z_{0\nu}|^2\int_0^l |\psi_\nu|^2 dx \end{aligned}\right\}. \tag{1115}$$

The corresponding integrals involving the derivatives of the normal functions are expressible directly in terms of these four by means of the relations (1060a), and need not be evaluated separately.

Using the expressions (1075) for the normal functions we find for the set (1115)

$$J_1 = 1 + \frac{\sinh(\rho_{S\nu} + \rho_{R\nu})\cdot\cosh(\rho_{S\nu} - \rho_{R\nu})}{\rho_{S\nu} + \rho_{R\nu}}, \tag{1115a}$$

$$J_2 = \frac{\sinh(\varepsilon_{S\nu} + \varepsilon_{R\nu})\cdot\cosh(\varepsilon_{S\nu} - \varepsilon_{R\nu})}{\varepsilon_{S\nu} + \varepsilon_{R\nu}} + \frac{\sin(\vartheta_{S\nu} + \vartheta_{R\nu})\cdot\cos(\vartheta_{S\nu} - \vartheta_{R\nu})}{\vartheta_{S\nu} + \vartheta_{R\nu}}, \tag{1115b}$$

$$J_3 = \frac{\sinh(\rho_{S\nu} + \rho_{R\nu})\cdot\cosh(\rho_{S\nu} - \rho_{R\nu})}{\rho_{S\nu} + \rho_{R\nu}} - 1, \tag{1115c}$$

$$J_4 = \frac{\sinh(\varepsilon_{S\nu} + \varepsilon_{R\nu})\cdot\cosh(\varepsilon_{S\nu} - \varepsilon_{R\nu})}{\varepsilon_{S\nu} + \varepsilon_{R\nu}} - \frac{\sin(\vartheta_{S\nu} + \vartheta_{R\nu})\cdot\cos(\vartheta_{S\nu} - \vartheta_{R\nu})}{\vartheta_{S\nu} + \vartheta_{R\nu}}. \tag{1115d}$$

In the special cases discussed above, these integrals simplify. For

example, in (1091), (1092), (1094), (1096), and (1099) for which λ_ν, $\rho_{S\nu}$, and $\rho_{R\nu}$ become pure imaginaries[1] we have

$$\left.\begin{aligned} J_1 = J_2 &= 1 + \frac{\sin(\vartheta_{S\nu} + \vartheta_{R\nu}) \cdot \cos(\vartheta_{S\nu} - \vartheta_{R\nu})}{\vartheta_{S\nu} + \vartheta_{R\nu}} \\ \text{and} \qquad J_3 &= -J_4 = J_{1,2} - 2 \end{aligned}\right\}. \tag{1116}$$

Particularly for (1094), for which the reflection coefficients are equal and opposite, $\vartheta_{S\nu} - \vartheta_{R\nu} = \pm\pi/2$, so that

$$J_1 = J_2 = J_4 = -J_3 = 1. \tag{1116a}$$

In (1093), which is the same as (1094) for real constant values of a_ν, we get

$$\left.\begin{aligned} J_1 &= -J_3 = 1 \\ J_2 &= J_4 = \frac{\sinh 2\varepsilon_S}{2\varepsilon_S} \end{aligned}\right\}. \tag{1116b}$$

Note that in (1094)

$$\int_0^l (\varphi_\mu^2 + q\psi_\mu^2)dx = J_1 + J_3 = 0,$$

as it should be, while for the application of (1082a) we have

$$\int_0^l (|\varphi_\mu|^2 + q|\psi_\mu|^2)dx = J_2 + J_4 = 2.$$

On the other hand, we have for (1095)

$$\int_0^l (|\varphi_\mu|^2 + q|\psi_\mu|^2)dx = J_2 - J_4 = \frac{2\sin(\vartheta_{S\nu} + \vartheta_{R\nu})\cos(\vartheta_{S\nu} - \vartheta_{R\nu})}{\vartheta_{S\nu} + \vartheta_{R\nu}}.$$

When $r_S r_R < 0$, $\vartheta_{S\nu} - \vartheta_{R\nu} = \pm\pi/2$, so that this is always zero. When $r_S r_R > 0$, $\vartheta_{S\nu} - \vartheta_{R\nu} = 0$, and by (1111a) we have

$$J_2 - J_4 = \frac{2\sin\pi\nu}{\pi\nu},$$

which vanishes except for $\nu = 0$. Except for this value, A_μ is given by (1082b) with

$$\int_0^l (\varphi_\mu^2 + q\psi_\mu^2)dx = J_1 - J_3 = 2.$$

8. Applications. Several simple cases will be chosen to illustrate the application of this method of analysis. Suppose we consider a nondissipative line with $Z_S = 0$ and $Z_R = \infty$ and a constant applied emf of

[1] In (1094) this excludes real constant values of a_ν, which are represented by (1093).

magnitude E. The switching instant is $t = 0$. Here $\lambda_\nu = j\omega_\nu$, while $r_S = -1$ and $r_R = +1$ so that $\rho_{S\nu} = j\vartheta_{S\nu} = -j\frac{\pi\nu}{2}$ and $\rho_{R\nu} = j\vartheta_{R\nu} = j\pi\nu$, which checks with (1111b). The latter also gives $\lambda_\nu = \frac{j\pi\nu}{2l}$ for ν-odd.

Substituting into (1075), the normal functions become

$$\varphi_\nu(x) = (-1)^{\frac{\nu+1}{2}}\sqrt{\frac{2}{l}}\sin\frac{\nu\pi}{2l}x,$$

$$\psi_\nu(x) = \frac{j(-1)^{\frac{\nu+1}{2}}}{Z_0}\sqrt{\frac{2}{l}}\cos\frac{\nu\pi}{2l}x.$$

Let us assume that no charge or current is present on the line initially. Then since $e_s(x, t) = E$ and $i_s(x, t) = 0$, (1077a) gives $P(x) = -E$ and $Q(x) = 0$. Using the results (1116a) in (1081) we have

$$A_\nu + A_{-\nu} = -\int_0^l E\varphi_\nu(x)dx$$
$$A_\nu - A_{-\nu} = 0,$$

from which we find

$$A_\nu = \frac{E}{\pi\nu}\sqrt{2l}\cdot(-1)^{\frac{\nu-1}{2}}.$$

Since in this case $p_\nu = v_0\lambda_\nu = j\frac{v_0\pi\nu}{2l}$, the transient voltage and current expressions evaluate to

$$\left.\begin{aligned} e_t(x,t) &= -\sum_{\nu=1,3,5\ldots}^{\infty}\frac{4E}{\nu\pi}\sin\frac{\nu\pi}{2l}x\cdot\cos\frac{v_0\nu\pi}{2l}t \\ i_t(x,t) &= \sum_{\nu=1,3,5\ldots}^{\infty}\frac{4E}{\nu\pi Z_0}\cos\frac{\nu\pi}{2l}x\cdot\sin\frac{v_0\nu\pi}{2l}t \end{aligned}\right\}. \tag{1117}$$

Making a simple trigonometric transformation, the net voltage and current expressions finally are

$$\left.\begin{aligned} e(x,t) &= E - \sum_{\nu=1,3,5\ldots}^{\infty}\frac{2E}{\nu\pi}\left\{\sin\frac{\nu\pi}{2l}(x - v_0t) + \sin\frac{\nu\pi}{2l}(x + v_0t)\right\} \\ i(x,t) &= -\sum_{\nu=1,3,5\ldots}^{\infty}\frac{2E}{\nu\pi Z_0}\left\{\sin\frac{\nu\pi}{2l}(x - v_0t) - \sin\frac{\nu\pi}{2l}(x + v_0t)\right\} \end{aligned}\right\}. \tag{1117a}$$

Each of these sums represents a pair of rectangular waves having a wave length equal to $4l$. The voltage waves have the amplitude $E/2$ and the current waves the amplitude $E/2Z_0$. The two waves in a pair travel in opposite directions with the velocity v_0, giving rise to a resulting picture which the student may sketch for himself as an exercise.

As a second example suppose we again consider a non-dissipative line

with an inductance L_s at the sending end and a capacitance C_r at the receiving end such that $L_s/C_r = Z_0^2$. This becomes case (1094) with

$$a_\nu = \frac{jL_s\omega_\nu}{L}$$

where L is the distributed inductance of the line. Since

$$r_{S_\nu} = \frac{jL_s\omega_\nu - L}{jL_s\omega_\nu + L}; \quad r_{R_\nu} = \frac{L - jL_s\omega_\nu}{L + jL_s\omega_\nu},$$

we have

$$\vartheta_{S_\nu} = \cot^{-1}\left(\frac{L_s\omega_\nu}{L}\right) = \vartheta_{R_\nu} + \frac{\pi}{2}.$$

The normal functions according to (1075) are

$$\varphi_\nu(x) = \sqrt{\frac{2}{l}} \cos\,(\omega_\nu x - \vartheta_{S_\nu}),$$

$$\psi_\nu(x) = \frac{1}{jZ_0}\sqrt{\frac{2}{l}} \sin\,(\omega_\nu x - \vartheta_{S_\nu}).$$

Let us again assume that a constant emf of E volts is suddenly applied at $t = 0$, and that the initial charge and current distributions are zero so that $P(x) = -E$ and $Q(x) = 0$. The coefficients for this case are given by (1082a). With (1116a) we have

$$A_\nu = -\frac{E}{2}\sqrt{\frac{2}{l}}\int_0^l \cos\,(\omega_\nu x - \vartheta_{S_\nu})\cdot dx,$$

which is easily evaluated to give

$$A_\nu = \frac{E\sqrt{l}}{4}\cdot\left(\frac{\sin\frac{\omega_\nu l}{2}}{\frac{\omega_\nu l}{2}}\right).$$

Thus the coefficients are real, and consequently even functions of ω_ν as they should be. The ω_ν's, according to (1113), are evidently odd functions of ν (opposite ν-values giving conjugate λ_ν-values). The sums in (1076) for the transient voltage and current are thus easily converted to extend only over positive values of ν. This is done merely for convenience. We then have for the net voltage and current

$$\left.\begin{aligned} e(x,t) &= E - \frac{E}{\sqrt{2}}\sum_{\nu=1}^{\infty}\left(\frac{\sin\frac{\omega_\nu l}{2}}{\frac{\omega_\nu l}{2}}\right)\cos\,(\omega_\nu x - \vartheta_{S_\nu})\cdot\cos v_0\omega_\nu t \\ i(x,t) &= -\frac{E}{Z_0\sqrt{2}}\sum_{\nu=1}^{\infty}\left(\frac{\sin\frac{\omega_\nu l}{2}}{\frac{\omega_\nu l}{2}}\right)\sin\,(\omega_\nu x - \vartheta_{S_\nu})\cdot\sin v_0\omega_\nu t \end{aligned}\right\}, \quad (1118)$$

which may readily be put into the wave form by applying the proper trigonometric formulae.

The ω_ν-values are easily found by plotting the relation (1113), which for this case is given by

$$\frac{L_S\omega_\nu}{L} = \cot\left(\frac{\omega_\nu l}{2} \pm \frac{\pi}{4}\right),$$

and reading off the ω-values at intersecting points.

The student should note that the series (1117) and (1117a) are harmonic and hence ordinary Fourier series, while those involved in (1118) are non-harmonic; i.e., the "frequencies" of the higher terms are not integer multiples of that of the first, as is evident from the determinantal equation. These are not Fourier series although the coefficients are determined in the Fourier manner.

This second example is a good illustration of how simple and direct this classical method of solution becomes when the process is properly organized. The result, of course, is in the form of an infinite series, but for practical computations it may be broken off after a relatively few terms. The result (1118) may easily be used to compute, for example, the voltage across the condenser at the receiving end by putting $x = l$. Utilizing the relations $\omega_\nu l = \vartheta_{S\nu} + \vartheta_{R\nu}$ and $\vartheta_{S\nu} = \vartheta_{R\nu} + \pi/2$, which hold for this case, we find for this voltage

$$e(l,t) = E\left\{1 - \sum_{\nu=1}^{\infty} \frac{1 + \sin\omega_\nu l - \cos\omega_\nu l}{2\omega_\nu l} \cdot \cos v_0\omega_\nu t\right\}. \tag{1119}$$

9. The properly terminated line. It was pointed out earlier that the cases corresponding to proper termination of the line become singular if treated in the above manner. This is due to the fact that this condition virtually corresponds to an infinitely long line for which the modes and eigenwerte become differentially spaced. The result for such a case may be evaluated by a subsequent limiting process. Thus if in (1117a), for example, we let $l \to \infty$, the sums go over into the familiar Fourier integrals

$$\left.\begin{aligned} e(x,t) &= E - \frac{E}{\pi}\int_0^\infty \frac{\sin\lambda(x - v_0t) + \sin\lambda(x + v_0t)}{\lambda}d\lambda \\ i(x,t) &= -\frac{E}{\pi Z_0}\int_0^\infty \frac{\sin\lambda(x - v_0t) - \sin\lambda(x + v_0t)}{\lambda}d\lambda \end{aligned}\right\}, \tag{1120}$$

where the continuous variable λ is the result of the limit

$$\frac{\nu\pi}{2l} \to \lambda$$

as $l \to \infty$ and $\nu \to \infty$. Since this limiting process has been described in detail in Chapter XI, we need not go into the matter any further here. This case, of course, is much more easily treated by other methods and is here merely of academic interest.

PROBLEMS TO CHAPTER XIII

13–1. Draw a series of wave pictures illustrating the solution (1117*a*) for the problem treated in section 8 of the text.

13–2. Make plots from which the ω_ν-values are determined for the second problem worked out in section 8 of the text. With these, evaluate a sufficient number of terms in the expression (1118) in order to obtain the general character of the traveling-wave picture.

13–3. Re-solve the problem referred to in 13–1 for the terminal conditions $Z_S = 0$, $Z_R = 0$.

13–4. For the non-dissipative line assume a rectangular charge distribution of length $l/4$ to be initially present in the center of the line (simplified picture of a charge induced by a lightning discharge). Determine the resulting wave picture as a function of time for the terminal conditions: (*a*) $Z_S = Z_R = 0$; (*b*) $Z_S = Z_R = \infty$; (*c*) $Z_S = 0$; $Z_R = \infty$.

13–5. Reconsider the problem referred to in 13–1, but take into account the resistance parameter of the line. Again interpret and plot the resulting wave picture.

13–6. A non-dissipative line has an inductance connected to its sending end, the far end being open. Find the response for a suddenly applied constant voltage for initial rest conditions.

13–7. Repeat Problem 13–6, replacing the inductance by a capacitance.

13–8. For a non-dissipative line $Z_S = 0$ and Z_R is given by (*a*) an inductance, (*b*) a capacitance. Find the response in each case for a suddenly applied constant voltage and initial rest conditions. Interpret and plot the resulting wave picture.

13–9. Repeat problem 13–8 for Z_R given by an inductance and capacitance in parallel, using Rayleigh's method for determining the coefficients A_ν and assuming initial rest conditions.

13–10. Reconsider Problem 13–4, taking into account the resistance of the line.

APPENDIX

LIST OF SYMBOLS

The more commonly used symbols are grouped at the head of the following list. The less common ones, or those appearing only locally in connection with more detailed treatments, are grouped under their respective chapter numbers. In some instances it occurs that the same letter or character is used to designate different quantities at different times. An attempt has been made to allow such multiplicity in the significance of a character only in detailed treatments which are unlikely to overlap. It is quite obvious that an attempt to avoid duplication altogether leads not only to an unwarrantedly complicated symbolism but to awkward and wholly unfamiliar forms in the mathematical results. The bracketed numbers following a character or group of characters refer to pages in the text where these are introduced and described, thus making their description in the present list unnecessary. In order to find the most pertinent definition, the symbol should be sought first under the chapter number in question.

Common Group

e, i, instantaneous voltage and current; E, I, complex amplitudes of voltage and current; ε, μ, dielectric constant and permeability; ϵ, error or tolerance; t_d, time delay.

λ, wave length; v, phase velocity; f, frequency; $\omega = 2\pi f$, angular frequency; $\tau = 1/f$, period; [3, 49].

R, G, [13], L, C, [23], distributed line parameters; x, l, [13, 41], distance along the line and line length.

E_g, Z_S, Z_R, [40, 162]; r_S, r_R, [42]; α, [37, 83]; Z_0, [40].

y_{11}, y_{22}, y_{12}, z_{11}, z_{22}, z_{12}, [135]; g_{11}, g_{22}, g_{12}, g_{21}, h_{11}, h_{22}, h_{12}, h_{21}, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, [137].

The complex quantities α, γ, γ_k, γ_{km}, $\cdots$ in rectangular form are written

$$\alpha_1 + j\alpha_2,\ \gamma_1 + j\gamma_2,\ \gamma_{1k} + j\gamma_{2k},\ \cdots \text{ etc.}$$

I

φ, $\dot{q}$, [13]; B_p, E_p, [15].

II

$E(x), I(x)$, [37]; E_S, E_R, y, z, [40]; ρ_S, ρ_R, [52]; k_S, k_R, [54]; ϑ_S, ϑ_R, [55]; Z_{oc}, Z_{sc}, [71].

III

$a, b, m, n, x, f_1(x), f_2(x)$, [84]; $\delta_\alpha, \delta_{\alpha 1}, \delta_{\alpha 2}$, [93]; λ_g, [102]; v_{ph}, v_g, [103, 105]; z_0, ζ, [107]; δ_{z_0}, [110]; $\gamma_E, \gamma_I, \gamma_{EI}, \gamma'$, [112].

IV

$E_1, E_2, I_1, I_2, B_{11}, D$, [134]; a, [151]; L_{11}, L_{22}, L_{12}, [153]; a', L_1, L_2, k, [154]; y_1, y_2, z_1, z_2, [158]; y_a, y_b, z_a, z_b, [159]; $z_A, z_B, z_C, y_A, y_B, y_C$, [157]; γ, [163, 178, 179]; Z_{01}, Z_{02}, r_S, r_R, [165]; θ, [169]; $Z_{I1}, Z_{I2}, r_{S1}, r_{S2}, r_{R1}, r_{R2}$, [171, 173]; Z_T, Y_π, Z_π, Z_I, [178, 179, 180].

V

$L_{ik}, S_{ik}, b_{ik}, B_{11}, D, Z_{11}, \lambda$, [184]; $b_{ik}^*, B_{11}^*, D^*, H$, [185]; A_k, [188]; $L_k, C_k, \omega_k, z_k, Y_{11}$, [189]; $\bar{B}_k, y_k$, [190]; $f(\lambda), h(\lambda), g(\lambda), k_k$, [192]; P, Q, [199]; R, [205]; R_{ik}, [208].

VI

$q_k, \dot{q}_k, i_k, \varphi_k, T$, [223]; e_{ck}, e_{rk}, V, F, [224]; $I_k, \bar{I}_k$, [226]; b_{ik}, [227]; X_{11}, [228]; α_{ks}, [229]; A, A', [231]; $\mathsf{L}, \mathsf{L}', \mathsf{R}, \mathsf{R}', \mathsf{S}, \mathsf{S}'$, [231, 232]; $k_{ik}, D(p), p_\nu, g_\nu$, [239]; $K_{ik}^{(\nu)}, l_{k\nu}, N_\nu$, [240]; α, [243]; β, [245]; $C_{rs}, G_{rs}, \Gamma_{rs}, d_{rs}$, [249]; $\mathsf{C}, \mathsf{G}, \mathsf{\Gamma}$, [251].

VII

$y_1, y_2, z_1, z_2, Z_T, \gamma$, [see IV]; δ_α, δ_z, [259]; Z_a, Z_b, [260]; y, z, [see II]; H, x, [263]; k, [267]; α, β, [268 or VI]; b_{ik}^*, D^*, [268]; B_{ik}^*, Z_{ik}^*, [269]; L_l, [271]; z', [273].

VIII

x, [291]; u_S, u_R, [292]; $f(x), R_t$, [295]; ρ, k, [296].

IX

x, ξ, η, [301]; z_{1k}, z_{2k}, x_k, R, [303]; W_{1k}, W_{2k}, [305]; $L_1, C_1, L_2, C_2, \omega_c, \omega_{c1}, \omega_{c2}, \omega_0, w$, [316-319]; $y_{1k}, y_{2k}, \gamma_k, y_k$, [325]; $m, z_{1km}', z_{2km}', \gamma_{km}$, [326]; x_{km}, [327]; W_{2km}, [331]; y_{1km}'', y_{2km}'', [335]; z_{1km}'', z_{2km}'', [336]; W_{1km}, [337]; $z_{1kmm'}', z_{2kmm'}', W_{1kmm'}, W_{2kmm'}$, etc., [343]; g, [345]; r, u, ρ, [353]; G, B, Y, Z, [356, 357]; $\mathcal{R}, \mathcal{X}$, [365]; $a_1, a_2, \cdots, \alpha_1, \alpha_2, \cdots$, [367]; $K(\cdots)$, [368].

X

Z_a, Z_b, γ, Z_0, [379]; z_a, z_b, z_0, [381]; y_0, [383]; γ^*, Z_0^*, [388]; H, ω_1, ω_2, $\cdots$, ω_a, ω_b, $\cdots$, ω_{-1}, ω_{-2}, $\cdots$, ω_{-a}, ω_{-b}, $\cdots$, ω_0, (Cauer), [389, 390]; w_1, w_2, $\cdots$, x, x_ν, [391, 392]; $F_1(x)$, $F_2(x)$, $\cdots$, [393]; $f(u)$, $F(u)$, k, [395]; φ, φ_ν, [396]; u, u_ν, snu, snu_ν, K, [397, 398]; ω_1, ω_2, $\cdots$, (Bode), [413]; P, Q, Q', [414]; ω_{pm}, [415]; δ_s, [416]; m_ι, [436]; α, β, [445 or VI or VII]; μ, λ', [445]; P, A, B, P_0, A_0, B_0, [446]; $\lambda = \gamma + j\omega$, [447].

XI

a_ν, p, [462]; δ, [463]; $g(\omega)$, [466]; $g_1(\omega)$, $g_2(\omega)$, $G(\omega)$, $\psi(\omega)$, [469]; α, [470]; $Y_{12}(\omega)$, $Z_R(\omega)$, $h(\omega)$, $H(\omega)$, $\theta(\omega)$, [475]; K, t_d, [476]; ω_1, [477]; $Si(x)$, [478]; τ_a, [479]; ω_ι, [481]; $\varphi = \varphi_1 + j\varphi_2$, $Ci(x)$, [482]; t', [485]; ω_{-1}, ω_1, ω_0, [486]; w, [487]; τ_{ph}, [491]; $A(\omega)$, A_0, v_a, t_0, [492]; k_ι, [500].

XII

γ, z_a, r_a, x_a, [509]; ρ_1, ρ_2, [511]; r_0, r, x, θ, [515]; a, b, ψ, [519]; γ_a, γ_b, [521]; δ, γ_m, y, [522]; δ', γ_m', [525]; h, x_1, x_2, x_3, g, ν, [529]; z_a', r_a', ρ, [535]; p, q, [543].

XIII

$\varphi_\nu(x)$, $\psi_\nu(x)$, p_ν, y_ν, z_ν, λ_ν, [557]; δ, σ, v_0, $Z_{0\nu}$, [558]; $Z_{S\nu}$, $Z_{R\nu}$, $r_{S\nu}$, $r_{R\nu}$, [559]; $\rho_{S\nu}$, $\rho_{R\nu}$, A_ν, [560]; $P(x)$, $Q(x)$, [561]; q, [563]; γ_ν, ω_ν, $\epsilon_{S\nu}$, $\epsilon_{R\nu}$, $\vartheta_{S\nu}$, $\vartheta_{R\nu}$, [570]; J_1, J_2, J_3, J_4, [572].

INDEX

A

B

C

O

P

Q

R